Mean-Field Pontryagin Maximum Principle

Mattia Bongini,[*] Massimo Fornasier,[†] Francesco Rossi,[‡] Francesco Solombrino[§]

April 10, 2015

Abstract

We derive a Maximum Principle for optimal control problems with constraints given by the coupling of a system of ODEs and a PDE of Vlasov-type. Such problems arise naturally as Γ-limits of optimal control problems subject to ODE constraints, modeling, for instance, external interventions on crowd dynamics. We obtain these first-order optimality conditions in the form of Hamiltonian flows in the Wasserstein space of probability measures with forward-backward boundary conditions with respect to the first and second marginals, respectively. In particular, we recover the equations and their solutions by means of a constructive procedure, which can be seen as the mean-field limit of the Pontryagin Maximum Principle applied to the discrete optimal control problems, under a suitable scaling of the adjoint variables.

Keywords: Sparse optimal control, mean-field limit, Γ-limit, optimal control with ODE-PDE constraints, subdifferential calculus, Hamiltonian flows.

1 Introduction

The study of large crowds of interacting agents has received a significantly growing attention in the mathematical literature of the last decade with applications in biology, ecology, social sciences, and economics. Starting from the seminal papers [25, 27, 37, 39], emphasis has been put on self-organization, i.e., the formation of macroscopic patterns from the superimposition of simple, reiterated, binary interactions. A quintessential situation is the convergence of a crowd to a common state, which may be called consensus, agreement, or rendezvous. Several examples show that spontaneous convergence to pattern formation is not always guaranteed, e.g., for highly dispersed initial configurations in consensus problems [17, 18, 21, 30], hence, the issue of controlling and stabilizing these systems arises naturally. Two major interpretations of control of multiagent systems have received much attention: on the one hand, with the *decentralized approach*, the problem is recast into a game-theoretic framework, where agents optimize their individual cost and solutions correspond to Nash equilibria. On the other hand, following the concept of *centralized intervention*, an external policy-maker controlling the dynamics is introduced, with the task of minimizing its intervention cost.

[*]Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany. `mattia.bongini@ma.tum.de`

[†]Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany. `massimo.fornasier@ma.tum.de`

[‡]Aix Marseille Université, CNRS, ENSAM, Université de Toulon, LSIS UMR 7296, 13397, Marseille, France. `francesco.rossi@lsis.org`

[§]Technische Universität München, Fakultät Mathematik, Boltzmannstrasse 3 D-85748, Garching bei München, Germany. `francesco.solombrino@ma.tum.de`

When dealing with large populations, in both cases one faces the well-known problem of the *curse of dimensionality*, term first coined by Bellman precisely in the context of dynamic optimization: the complexity of numerical computations of the solutions of the above problems blows up as the size of the population increases. A possible way out is the so-called *mean-field approach*, where the individual influence of the entire population on the dynamics of a single agent is replaced by an averaged one. This substitution principle results in a unique mean-field equation and allows the computation of solutions, cutting loose from the dimensionality.

In the game-theoretic setting, the mean-field approach has led to the development of *mean-field games* [26, 29], which model populations whose agents are competing *freely* with the others towards the maximization of their individual payoff, as for instance in the financial market. The landmark feature of such systems is their capability to autonomously stabilize without external intervention. However, in reality, societies exhibit either convergence to undesired patterns or tendencies toward instability that only an external government can successfully dominate. The need of such interventions, together with the limited amount of resources that governments have at their disposal, makes the design of parsimonious stabilization strategies a key issue, which has been extensively studied in the context of dynamics given by systems of ODEs, see [7, 8, 9, 10, 13].

Nevertheless, the concept of sparse control has to be handled with care when trying to generalize it at the level of a mean-field dynamics. Indeed, the indistinguishability of agents is a fundamental property of the mean-field setting, and it is in sharp contrast with controls acting sparsely on specific agents. Figuratively, trying to stabilize a huge crowds with these controls is like steering a river by means of toothpicks! A first solution to this ambiguity was given in [6, 23], where the control is defined as a locally Lipschitz feedback control with respect to the state variables, and sparsity refers to its property of having a small support. Such concept was successfully used in [33] to implement sparse stabilizers for a consensus problem. This interpretation of sparsity appears also in the framework of the control of more classical PDEs, see [16, 34, 35, 38]. An alternative solution for a proper definition of sparse mean-field control was proposed in [22], where the control is sparsely applied on a finite number of individuals immersed in the mean-field dynamics of the rest of the population, resulting in a system where the controlled ODEs are coupled with a control-free mean-field PDE (but indirectly controlled via the coupling). The same kind of control was considered in [1] to model the efficient evacuation of a large crowd of pedestrians with the help of very few informed agents.

While in the context of mean-field games and optimal control problems with PDE constraints, first-order optimality conditions have received enormous attention, see for instance [5, 11, 14, 36], up to now no corresponding results have appeared in the literature for coupled ODE-PDE systems of the kind considered in [22], to the best of our knowledge. This paper is devoted to the development of a Pontryagin Maximum Principle to characterize optima of such control problems. We first remark that we are not interested in all possible optima, but mainly on those which arise as limits of optimal strategies of the original discrete problems. We call this subclass of the set of optima *mean-field optimal controls* (see Definition 1.4). The interest in this class complies with the wish of using the continuous models as approximations of the finite-dimensional ones. Furthermore in the model cases considered in [22, 23], it is exactly the existence of mean-field optimal controls that is proved.

We summarize our result, borrowing a leaf from the diagram in [14], as follows:
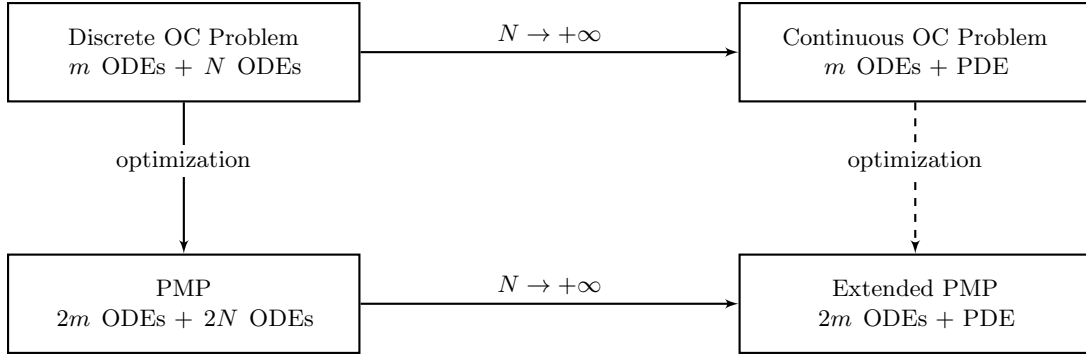
We shall provide a set of hypotheses for which the dashed line from the upper-right to the bottom-right box is valid. Our strategy shall be the following: we apply the Pontryagin Maximum Principle (see e.g. [15, Theorem 23.11]) to the finite-dimensional optimal control problems (the solid line from the upper-left to the bottom-left box), and we pass to the mean-field limit the system of equations obtained with this procedure (the solid line from the bottom-left to the bottom-right box). The derived limit equation for the state and the (rescaled) adjoint variables are obtained in the form of Hamiltonian flows in the Wasserstein space of probability measures, in the sense of [3]. The result will be a first-order condition valid for all *mean-field optimal controls*. The existence of such controls is also proved (see Corollary 2.15).

More formally, we are interested in deriving optimality conditions for the solutions of the following optimal control problem subject to coupled ODE-PDE constraints.

Problem 1. For $T > 0$ fixed, find $u^* \in L^1([0,T]; \mathcal{U})$ minimizing the cost functional

$$F(u) = \int_0^T \left[L(y(t), \mu(t)) + \gamma(u(t)) \right] dt, \qquad (1.1)$$

where (y, μ) solve

$$\begin{cases} \dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m, \\ \partial_t \mu(t) = -\nabla_x \cdot \left[(K \star \mu(t) + g(y(t))) \mu(t) \right], \end{cases} \qquad (1.2)$$

for the given initial datum $(y(0), \mu(0)) = (y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$.

Here, γ is a strictly convex cost functional, the finite dimensional set of controls $\mathcal{U}$ is convex and compact, B_k are constant matrices, and $\mathcal{P}_c(\mathbb{R}^d)$ is the set of probability measures on $\mathbb{R}^d$ with compact support.

We shall prove the following main result.

Theorem 1.1. *Fix an initial datum $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ and assume that Hypotheses (H) in Section 1.1 hold. Then there exists a mean-field optimal control for Problem 1. Furthermore, if u^* is a mean-field optimal control for Problem 1 and (y^*, μ^*) is the corresponding trajectory, then (u^*, y^*, μ^*) satisfies the following* **extended Pontryagin Maximum Principle***:*

There exists $(q^*(\cdot), \nu^*(\cdot)) \in \mathrm{Lip}([0,T]; \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^{2d}))$ such that

- *there exists $R_T > 0$, depending only on $y^0, \mathrm{supp}(\mu^0), d, K, g, f_k, B_k, \mathcal{U},$ and T, such that $\mathrm{supp}(\nu^*(\cdot)) \subseteq B(0, R_T)$ and it satisfies $\nu^*(t)(E \times \mathbb{R}^d) = \mu^*(t)(E)$ for all $t \in [0,T]$ and for every Borel set $E \subseteq \mathbb{R}^d$;*

- *it holds*

$$
\begin{cases}
\dot{y}_k^* & = \nabla_{q_k} \mathbb{H}_c(y^*, q^*, \nu^*, u^*), \\
\dot{q}_k^* & = -\nabla_{y_k} \mathbb{H}_c(y^*, q^*, \nu^*, u^*), \\
\partial_t \nu^* & = -\nabla_{(x,r)} \cdot ((J \nabla_\nu \mathbb{H}_c(y^*, q^*, \nu^*, u^*)) \nu^*), \\
u^* & = \arg\max_{u \in \mathcal{U}} \mathbb{H}_c(y^*, q^*, \nu^*, u)
\end{cases}
\tag{1.3}
$$

where $J \in \mathbb{R}^{2d \times 2d}$ is the symplectic matrix

$$
J = \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix},
$$

the Hamiltonian $\mathbb{H}_c : \mathbb{R}^{2dm} \times \mathcal{P}_c(\mathbb{R}^{2d}) \times \mathbb{R}^D \to \mathbb{R}$ is defined as

$$
\mathbb{H}_c(y, q, \nu, u) = \begin{cases} \mathbb{H}(y, q, \nu, u) & \text{if } \mathrm{supp}(\nu) \subseteq \overline{B(0, R_T)}, \\ +\infty & \text{elsewhere;} \end{cases}
$$

and $\mathbb{H} : \mathbb{R}^{2dm} \times \mathcal{P}_c(\mathbb{R}^{2d}) \times \mathbb{R}^D \to \mathbb{R}$ is defined as

$$
\mathbb{H}(y, q, \nu, u) = \frac{1}{2} \int_{\mathbb{R}^{4d}} (r - r') \cdot K(x - x') \, d\nu(x, r) \, d\nu(x', r') + \int_{\mathbb{R}^{2d}} r \cdot g(y)(x) d\nu(x, r)
$$
$$
+ \sum_{k=1}^m \int_{\mathbb{R}^{2d}} q_k \cdot K(y_k - x) \, d\nu(x, r) + \sum_{k=1}^m q_k \cdot (f_k(y) + B_k u) - L(y, \pi_{1\#} \nu) - \gamma(u).
$$
$$
\tag{1.4}
$$

- *the following conditions for system (1.3) hold at time 0: $y^*(0) = y^0$ and $\nu^*(0)(E \times \mathbb{R}^d) = \mu^0(E)$ for every Borel set $E \subseteq \mathbb{R}^d$,*

- *the following conditions for system (1.3) hold at time T: $q^*(T) = 0$ and $\nu^*(T)(\mathbb{R}^d \times E) = \delta_0(E)$ for every Borel set $E \subseteq \mathbb{R}^d$, where δ_0 is the Dirac measure centered in 0.*

As already mentioned, the formulation given above shows that the dynamics of (y^*, q^*, ν^*) is essentially an Hamiltonian flow in the Wasserstein space of probability measures with respect to state and adjoint variables with Hamiltonian $\mathbb{H}$, in the sense of [3]. The definition of $\mathbb{H}_c$ is introduced to simplify some technical details and does not alter the result. This fact is remarkably consistent with the dynamics (1.2), since both are flows in a Wasserstein space. We believe that this formulation of the optimality conditions making use of the formalism of subdifferential calculus in Wasserstein spaces of probability measures constitutes one of the novelties of the work.

Remark 1.2. For every (y, q, ν) with $\text{supp}(\nu) \subseteq \overline{B(0, R_T)}$, (1.4) immediately implies that

$$\overline{u} \in \arg \max_{u \in \mathcal{U}} \mathbb{H}_c(y, q, \nu, u) \iff \overline{u} \in \arg \max_{u \in \mathcal{U}} \left(\sum_{k=1}^{m} q_k \cdot B_k u - \gamma(u) \right).$$

Then, the strict convexity of γ and the convexity and the compactness of $\mathcal{U}$ imply that $\overline{u}$ is uniquely determined by (y, q, ν). This is the reason why we write the equality symbol in $u^* = \arg \max_{u \in \mathcal{U}} \mathbb{H}_c(y^*, q^*, \nu^*, u)$ in place of an inclusion.

We point out the difference between the usual gradient in $\mathbb{R}^{2d}$ with respect to the state variables x and the adjoint variables r, denoted by $\nabla_{(x,r)}$, and the Wasserstein gradient ∇_ν of $\mathbb{H}_c$, which, as shown in Section 4, whenever ν has supported contained in $B(0, R_T)$ can be computed explicitly as follows:

- For $l = 1, \ldots, d$, it holds

$$
\begin{aligned}
\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r) \cdot e_l = & \int_{\mathbb{R}^{2d}} (r - r') \cdot (\mathbf{D}K(x - x')e_l) \, d\nu(x', r') + r \cdot (\mathbf{D}_x g(y)(x)e_l) \\
& - \sum_{k=1}^{m} q_k \cdot (\mathbf{D}K(y_k - x)e_l) - \nabla_\xi \ell(y, x, \int \omega \mu) \cdot e_l \\
& - \left(\nabla_\varsigma \ell(y, x, \int \omega \mu) \mathbf{D}\omega(x) \right) \cdot e_l.
\end{aligned}
\tag{1.5}
$$

  These are the components of $\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r)$ in the x_l coordinates.

- For $l = d + 1, \ldots, 2d$ it holds

$$\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r) \cdot e_l = \int_{\mathbb{R}^{2d}} K(x - x') \cdot e_{l-d} \, d\nu(x', r') + g(y)(x) \cdot e_{l-d}. \tag{1.6}$$

  These are the components of $\nabla_\nu \mathbb{H}_c(y, q, \nu, u)(x, r)$ in the r_{l-d} coordinates.

In (1.5) and (1.6), the functions $\ell \in \mathcal{C}^2(\mathbb{R}^{dm} \times \mathbb{R}^d \times \mathbb{R}^d; \mathbb{R})$ and $\omega \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d)$ are related to the functional L in (1.1) via

$$L(y, \mu) = \int_{\mathbb{R}^d} \ell\left(y, x, \int \omega \mu\right) d\mu(x),$$

where $\int \omega \mu := \omega \mu(\mathbb{R}^d)$, while $\nabla_\xi \ell$ and $\nabla_\varsigma \ell$ denote the partial derivatives of the function $\ell(\eta, \xi, \varsigma)$, and $\mathbf{D}\omega(x)$ is the Jacobian of the function ω evaluated at x. Notice that $\nabla_\nu \mathbb{H}(y, q, \nu, u)$ actually does not depend on u, as a consequence of the fact that the control does not act directly on the PDE component of (1.2).

The main tool we use to prove Theorem 1.1 is the Pontryagin Maximum Principle (henceforth, simply addressed as PMP) for optimal control problems with ODE constraint. We shall apply it to the following finite-dimensional problems, whose constraints converge to the coupled ODE-PDE system of Problem 1, as we will show in Section 2. For this reason, we call Theorem 1.1 the *extended PMP*.

Problem 2. For $T > 0$ fixed, find $u^* \in L^1([0, T]; \mathcal{U})$ minimizing the cost functional

$$F_N(u) = \int_0^T \left[L(y(t), \mu_N(t)) + \gamma(u(t)) \right] dt, \tag{1.7}$$

---

where (y, μ_N) solve

$$\begin{cases} \dot{y}_k = \frac{1}{N} \sum_{j=1}^{N} K(y_k - x_j) + f_k(y) + B_k u, & k = 1, \ldots, m \\ \dot{x}_i = \frac{1}{N} \sum_{j=1}^{N} K(x_i - x_j) + g(y)(x_i), & i = 1, \ldots, N, \end{cases} \qquad (1.8)$$

for the given initial datum $(y(0), x(0)) = (y^0, x^0) \in \mathbb{R}^{dm} \times \mathbb{R}^{dN}$, where

$$\mu_N(t)(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i(t)),$$

is the empirical measure centered on the trajectory $x(\cdot) = (x_1(\cdot), \ldots, x_N(\cdot))$.

The extended PMP will be derived after reformulating the finite-dimensional PMP applied to Problem 2 in terms of the empirical measure in the product space of state variables x_i and adjoint variables p_i, defined as

$$\nu_N(x, r) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i, r - N p_i).$$

Notice that rescaling the adjoint variables p_i by the number N of agents is needed in order to observe a nontrivial dynamics in the limit (see also Remark 3.7); indeed, within this scaling, the right-hand side of the finite-dimensional PMP is brought back to the form considered, for instance, in [19], with a different Hamiltonian.

The following diagram recollects the strategy of the proof, making use of the notation already introduced and reporting in which part of the paper each result is proved:
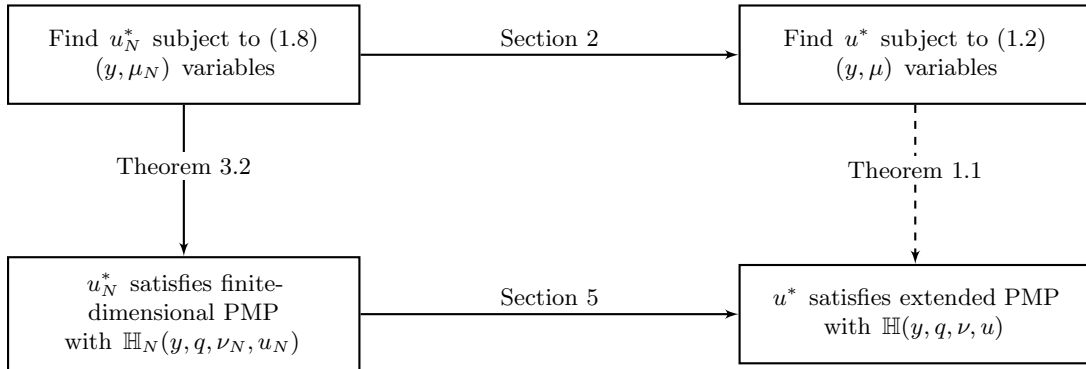
The structure of the paper is the following. In Section 1.1 we recall notations and the main Hypotheses (H). In Section 2, we study the controlled dynamics subject to a coupled ODE-PDE constraint of the form (1.2), establishing existence and uniqueness results for solutions. In Section 3 we study the finite-dimensional Problem 2, and apply the PMP to it. In Section 4, we recall basic facts about subdifferential calculus in Wasserstein spaces, and we explicitly compute $\nabla_\nu \mathbb{H}_c$. In Section 5, we prove the extended PMP, i.e., Theorem 1.1. Finally, Section 6 is devoted to the study of an interesting example of Problem 1, the Cucker-Smale system.

1.1 Notation and Hypotheses (H)

We start this section by recalling the notation used throughout the paper.

The constants d, D are two positive integers (the dimension of the space of the agents and of the control, respectively), $T > 0$ (the end time of the optimization procedure), and $\mathcal{U}$ is a *convex compact* subset of $\mathbb{R}^D$ (set in which controls take values).

Functionals have the following expressions: $K : \mathbb{R}^d \to \mathbb{R}^d$, each f_k satisfies $f_k : \mathbb{R}^{dm} \to \mathbb{R}^d$, and for every $y \in \mathbb{R}^{dm}$ and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, $g(y) : \mathbb{R}^d \to \mathbb{R}^d$ and $L(y, \mu) : \mathbb{R}^d \to \mathbb{R}$. The matrices B_k are constant $d \times D$ matrices.

The space $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures which take values on $\mathbb{R}^n$, while the space[1] $\mathcal{P}_p(\mathbb{R}^n)$ is the subset of $\mathcal{P}(\mathbb{R}^n)$ whose elements have finite p-th moment, i.e.,

$$\int_{\mathbb{R}^n} \|x\|^p d\mu(x) < +\infty.$$

We denote by $\mathcal{P}_c(\mathbb{R}^n)$ the subset of $\mathcal{P}_1(\mathbb{R}^n)$ which consists of all probability measures with compact support. Notice that, if $(\mu_n)_{n \in \mathbb{N}}$ is a sequence in $\mathcal{P}_c(\mathbb{R}^n)$ and it exists $R > 0$ such that $\mathrm{supp}(\mu_n) \subseteq B(0, R)$ for all $n \in \mathbb{N}$, then $(\mu_n)_{n \in \mathbb{N}}$ is compact in $\mathcal{P}_p(\mathbb{R}^n)$ for all $p \geq 1$.

For any $\mu \in \mathcal{P}(\mathbb{R}^n)$ and any Borel function $r : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, we denote by $r_\# \mu \in \mathcal{P}(\mathbb{R}^{n_2})$ the *push-forward of μ through r*, defined by

$$r_\# \mu(B) := \mu(r^{-1}(B)) \quad \text{for every Borel set } B \text{ of } \mathbb{R}^{n_2}.$$

In particular, if one considers the projection operators π_1 and π_2 defined on the product space $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, for every $\rho \in \mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ we call *first* (resp., *second*) *marginal* of ρ the probability measure $\pi_{1\#}\rho$ (resp., $\pi_{2\#}\rho$). Given $\mu \in \mathcal{P}(\mathbb{R}^{n_1})$ and $\nu \in \mathcal{P}(\mathbb{R}^{n_2})$, we denote with $\Gamma(\mu, \nu)$ the subset of all probability measures in $\mathcal{P}(\mathbb{R}^{n_1} \times \mathbb{R}^{n_2})$ with first marginal μ and second marginal ν.

On the set $\mathcal{P}_p(\mathbb{R}^n)$ we shall consider the following distance, called the *Wasserstein or Monge-Kantorovich-Rubinstein distance*,

$$\mathcal{W}_p^p(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^{2n}} \|x - y\|^p d\rho(x, y) : \rho \in \Gamma(\mu, \nu) \right\}. \tag{1.9}$$

If $p = 1$ we have the following equivalent expression for the Wasserstein distance:

$$\mathcal{W}_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^n} \varphi(x) d(\mu - \nu)(x) : \varphi \in \mathrm{Lip}(\mathbb{R}^n), \ \mathrm{Lip}(\varphi) \leq 1 \right\}.$$

We denote by $\Gamma_o(\mu, \nu)$ the set of optimal plans for which the minimum is attained, i.e.,

$$\rho \in \Gamma_o(\mu, \nu) \iff \rho \in \Gamma(\mu, \nu) \text{ and } \int_{\mathbb{R}^{2n}} \|x - y\|^p d\rho(x, y) = \mathcal{W}_p^p(\mu, \nu).$$

It is well-known that $\Gamma_o(\mu, \nu)$ is non-empty for every $(\mu, \nu) \in \mathcal{P}_p(\mathbb{R}^n) \times \mathcal{P}_p(\mathbb{R}^n)$, hence the infimum in (1.9) is actually a minimum. For more details, see e.g. [40, 4].

For any $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and $K : \mathbb{R}^d \to \mathbb{R}^d$, the notation $K \star \mu$ stands for the convolution of K and μ, i.e.,

$$(K \star \mu)(x) = \int_{\mathbb{R}^d} K(x - x') d\mu(x');$$

this quantity is well-defined whenever K is continuous and *sublinear*, i.e., there exists C such that $\|K(\xi)\| \leq C(1 + \|\xi\|)$ for all $\xi \in \mathbb{R}^d$. Furthermore we shall deal also with the convolution $(\nabla_{(x', r')}\langle r', K(x')\rangle) \star \nu$ in $\mathbb{R}^{2d}$, whose explicit expression is

$$\left((\nabla_{(x', r')}\langle r', K(x')\rangle) \star \nu \right)(x, r) = \int_{\mathbb{R}^{2d}} \left(\nabla_{(x', r')}\langle r - r', K(x - x')\rangle \right) d\nu(x', r').$$

---

[1] We follow the notation of [4].

Notice that, under the hypotheses we are going to make, this convolution is not always well-defined for $\nu \in \mathcal{P}_1(\mathbb{R}^{2d})$. It is nonetheless well-defined for measures $\nu \in \mathcal{P}_c(\mathbb{R}^{2d})$, that is to say for all the cases that will appear in the sequel.

We shall denote with $\mathcal{M}_b(\mathbb{R}^{n_1}; \mathbb{R}^{n_2})$ the space of bounded Radon vector measures from $\mathbb{R}^{n_1}$ to $\mathbb{R}^{n_2}$, and with $\| \cdot \|_{\mathcal{M}_b(\mathbb{R}^{n_1}; \mathbb{R}^{n_2})}$ the total variation norm on it. If $\omega \in \mathcal{C}(\mathbb{R}^d; \mathbb{R}^d)$ is sublinear and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, the Radon measure $\omega\mu \in \mathcal{M}_b(\mathbb{R}^d; \mathbb{R}^d)$ is defined as

$$\omega\mu(E) := \int_E \omega(x)d\mu(x), \quad \text{for every } E \subset \mathbb{R}^d \text{ bounded.}$$

We shall denote with $\int \omega\mu := \omega\mu(\mathbb{R}^d)$.

In what follows, we shall consider the space $\mathcal{X} := \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d)$, together with the following distance

$$\|(y, \mu) - (y', \mu')\|_{\mathcal{X}} := \|y - y'\| + \mathcal{W}_1(\mu, \mu'), \tag{1.10}$$

where $\|y - y'\| := \sum_{k=1}^m \|y_k - y'_k\|_{\ell_2(\mathbb{R}^d)}$.

Finally, for every $N \in \mathbb{N}$, the mapping $\Pi_N : \mathbb{R}^{2dN} \to \mathcal{P}_1(\mathbb{R}^{2d})$ is defined as follows

$$\Pi_N : (x_1, p_1, \ldots, x_N, p_N) \mapsto \frac{1}{N} \sum_{i=1}^N \delta(\cdot - x_i, \cdot - Np_i). \tag{1.11}$$

Henceforth, we assume that the following regularity properties hold.

Hypotheses (H)

(K) The function $K \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d)$ is odd and sublinear, i.e., there exists $C_K > 0$ such that for all $x \in \mathbb{R}^d$ it holds
$$\|K(x)\| < C_K(1 + \|x\|).$$

(L) The function $L : \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ is
$$L(y, \mu) = \int_{\mathbb{R}^d} \ell\left(y, x, \int \omega\mu\right) d\mu(x),$$
with $\ell \in \mathcal{C}^2(\mathbb{R}^{dm} \times \mathbb{R}^d \times \mathbb{R}^d; \mathbb{R})$ and $\omega \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d)$.

(G) The function $g \in \mathcal{C}^2(\mathbb{R}^{dm}; \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d))$ satisfies for all $x \in \mathbb{R}^d$ and all $y \in \mathbb{R}^{dm}$
$$g(y)(x) \cdot x \leq G_1\|x\|^2 + G_2 \max_{l=1,\ldots,m} \|y_l\|^2 + G_3,$$
where the constants G_1, G_2 and G_3 are independent on x and y.

(F) For each $k = 1, \ldots, m$, the function $f_k \in \mathcal{C}^2(\mathbb{R}^{dm}; \mathbb{R}^d)$ satisfies for all $y \in \mathbb{R}^{dm}$
$$f_k(y) \cdot y_k \leq F_1 \max_{l=1,\ldots,m} \|y_l\|^2 + F_2,$$
where the constants F_1 and F_2 are independent on y and k.

(U) The set $\mathcal{U} \subseteq \mathbb{R}^D$ is compact and convex.

(γ) The function $\gamma : \mathcal{U} \to \mathbb{R}$ is strictly convex.

Remark 1.3. We briefly compare Hypotheses (H) with those of [5, 11]. In [5], which deals with an SDE-constrained optimal control problem, $\mathcal{C}^{1,1}$ functionals with respect to state variables and the control are considered. Therefore our hypotheses are just slightly more restrictive. On the other hand, we do not require differentiability of the running cost. The authors of [11] deal, instead, with a mean-field game type optimality conditions to model evacuation scenarios. They derive a first-order condition under the hypotheses of continuous differentiability of the functionals with respect to the state variables together with convexity and positivity assumptions. Furthermore, they deal specifically with an L^2 control cost, while we allow ours to be strictly convex.

We now give the rigorous definition of *mean-field optimal control*.

Definition 1.4. Let $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ be given. An optimal control u^* for Problem 1 with initial datum (y^0, μ^0) is a *mean-field optimal control* if there exists a sequence $(u_N)_{N \in \mathbb{N}} \subset L^1([0,T]; \mathcal{U})$ and a sequence $(\mu_N^0)_{N \in \mathbb{N}} \in \mathcal{P}_c(\mathbb{R}^d)$ such that

(i) for every $N \in \mathbb{N}$, $\mu_N^0(\cdot) := \frac{1}{N} \sum_{i=1}^N (\cdot - x_{i,N}^0)$ is a sequence of empirical measures for some $x_{i,N}^0 \in \text{supp}(\mu^0) + \overline{B(0,1)}$ such that $\mu_N^0 \rightharpoonup \mu^0$ weakly* in the sense of measures;

(ii) for every $N \in \mathbb{N}$, u_N^* is a solution of Problem 2 with initial datum (y^0, μ_N^0);

(iii) there exists a subsequence of $(u_N)_{N \in \mathbb{N}}$ converging weakly in $L^1([0,T]; \mathcal{U})$ to u^*.

Remark 1.5. As mentioned before, the above definition is motivated by our interest in optimizers that are close to optimal controls for the original finite-dimensional problems. Notice also, that since the measures μ_N^0 have all compact support contained in $\text{supp}(\mu^0) + \overline{B(0,1)}$, the build a compact sequence in $\mathcal{P}_p(\mathbb{R}^n)$ for all $p \geq 1$, and therefore, due to weak* convergence to μ^0, we also have that $\lim_{N \to \infty} \mathcal{W}_p(\mu_N^0, \mu^0) = 0$.

2 The coupled ODE-PDE dynamics

In this section, we first recall results for PDE equations of transport type with nonlocal interaction velocities, like the one appearing in the second equation of (1.2). We then study the coupled ODE-PDE dynamics (1.2) and we state existence and uniqueness results of solutions, together with continuous dependence on the initial data (y^0, μ^0) and on the control u. The proofs follow closely in the footsteps of similar results in [3, 22, 31, 32]. We also show that finite-dimensional ODE dynamics (1.8) are embedded in (1.2), in the sense that the solution of (1.2) with an initial data that is an empirical measure coincides with the empirical measure with support on the solution of (1.8).

2.1 Transport PDE equations with nonlocal interaction

In this section, we study equations for the dynamics of measures, recalling results of existence and uniqueness. We first define the meaning of solution for the equation

$$\partial_t \mu(t) = -\nabla_x \cdot (v(t, x, \mu(t))\mu(t)), \tag{2.1}$$

where $v : [0,T] \times \mathbb{R}^n \times \mathcal{P}_1(\mathbb{R}^n) \to \mathbb{R}^n$ is a given vector field and $n \in \mathbb{N}$ is the dimension of the underlying Euclidean space.

Definition 2.1. We say that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^n)$ is a solution of (2.1) if the following holds:

(i) μ has uniformly compact support, i.e., there exists $R > 0$ such that $\text{supp}(\mu(\cdot)) \in B(0,R)$;

(ii) μ is continuous with respect to the Wasserstein distance $\mathcal{W}_1$;

(iii) μ satisfies (2.1) in the weak sense, i.e. (see [4, Equation (8.1.4)]),

$$\frac{d}{dt} \int_{\mathbb{R}^n} \phi(x)\, d\mu(t)(x) = \int_{\mathbb{R}^n} \nabla\phi(x) \cdot v(t, x, \mu(t))\, d\mu(t)(x),$$

for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^n; \mathbb{R})$.

Now, we can formally define the concept of solution of the controlled ODE-PDE system (1.2), which applies, *mutatis mutandis*, to system (1.3) as well.

Definition 2.2. Let $u \in L^1([0,T]; \mathcal{U})$ and $(y^0, \mu^0) \in \mathcal{X}$, with μ^0 of bounded support, be given. We say that a map $(y, \mu) : [0,T] \to \mathcal{X}$ is a solution of the system (1.2) with control u if

(i) $(y(0), \mu(0)) = (y^0, \mu^0)$;

(ii) the solution is continuous in time with respect to the metric (1.10) in $\mathcal{X}$;

(iii) the y coordinates define a Carathéodory solution of the following controlled ODE problem

$$\dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m,$$

for all $t \in [0,T]$;

(iv) μ is a solution of (2.1), where $v : [0,T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}^d$ is the time-varying vector field defined as follows
$$v(t, x, \mu(t))(x) := (K \star \mu(t) + g(y(t)))(x).$$

We now derive the existence of solutions of (1.2) as limits for $N \to \infty$ of the system of ODE (1.8). We first prove that solutions of (1.8) coincide with specific solutions of (1.2). We then prove the limit result with the help of Lemmata 2.4 and 2.5.

Proposition 2.3. *Let N be fixed, and the control $u \in L^1([0,T]; \mathcal{U})$ be given. Let $(y, x_N) : [0,T] \to \mathcal{X}$ be the corresponding solution of (1.8), with $x_N(t) = (x_{1,N}(t), \ldots, x_{N,N}(t))$. Then, the couple $(y, \mu_N) : [0,T] \to \mathbb{R}^{dm+dN}$, with $\mu_N(t)$ being the empirical measure*

$$\mu_N(t)(x) := \frac{1}{N} \sum_{i=1}^N (x - x_{i,N}(t)),$$

is a solution of (1.2) with control u.

Proof. It can be easily proved by rewriting (1.2) with μ_N and arguing exactly as in [23, Lemma 4.3]. □

Lemma 2.4. *Let $K : \mathbb{R}^d \to \mathbb{R}^d$ satisfy (K) and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$. Then for all $y \in \mathbb{R}^d$ it holds*

$$\|(K \star \mu)(y)\| \le C_K \left(1 + \|y\| + \int_{\mathbb{R}^d} \|x\| d\mu(x) \right).$$

Proof. See, for instance, [23, Lemma 6.4]. □

Lemma 2.5. *Let $K : \mathbb{R}^d \to \mathbb{R}^d$ satisfy (K) and let $\mu^1 : [0, T] \to \mathcal{P}_c(\mathbb{R}^d)$ and $\mu^2 : [0, T] \to \mathcal{P}_1(\mathbb{R}^d)$ be two continuous maps with respect to $\mathcal{W}_1$ satisfying*

$$\mathrm{supp}(\mu^1(t)) \cup \mathrm{supp}(\mu^2(t)) \subseteq B(0, R),$$

for every $t \in [0, T]$, for some $R > 0$. Then for every $\rho > 0$ there exists constant $L_{\rho,R}$ such that

$$\|K \star \mu^1(t) - K \star \mu^2(t)\|_{L^\infty(B(0,\rho))} \leq L_{\rho,R} \mathcal{W}_1(\mu^1(t), \mu^2(t))$$

for every $t \in [0, T]$.

Proof. A proof of this result may be found, for instance, in [23, Lemma 6.7]. □

Proposition 2.6. *Let $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, and $\mu_N^0(\cdot)$ be as in Definition 1.4–(i). Let $(u_N)_{N\in\mathbb{N}} \subseteq L^1([0, T]; \mathcal{U})$ be a sequence of controls such that $u_N \rightharpoonup u$, for some $u \in L^1([0, T]; \mathcal{U})$.*

Then, the sequence of solutions $(y_N, \mu_N) \in \mathrm{Lip}([0, T]; \mathcal{X})$ of (1.8) with initial data (y^0, μ_N^0) and control u_N converges to a solution $(y, \mu) \in \mathrm{Lip}([0, T]; \mathcal{X})$ of (1.2) with initial data (y^0, μ^0) and control u. Moreover, there exists $\rho_T > 0$, depending only on $y^0, \mathrm{supp}(\mu^0), K, g, f_k, B_k, \mathcal{U}$, and T, such that for every $N \in \mathbb{N}$, for every $k = 1, \ldots, m$ and for every $t \in [0, T]$ it holds

$$\|y_{k,N}(t)\|, \|y_k(t)\| \leq \rho_T \quad and \quad \mathrm{supp}(\mu_N(t)), \mathrm{supp}(\mu(t)) \subseteq B(0, \rho_T).$$

Proof. We start by fixing $N > 0$ and estimating the growth of $\|y_{k,N}(t)\|^2 + \|x_{i,N}(t)\|^2$ for $k = 1, \ldots, m$ and $i = 1, \ldots N$. Let $\Sigma = \{(l, j) : l = 1, \ldots, m$ and $j = 1, \ldots N\}$. From Hypotheses (H), Lemma 2.4 and the compactness of $\mathcal{U}$, it holds

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\left(\|y_{k,N}\|^2 + \|x_{i,N}\|^2\right) &= \dot{y}_{k,N} \cdot y_{k,N} + \dot{x}_{i,N} \cdot x_{i,N} \\
&= ((K \star \mu_N)(y_{k,N}) + f_k(y) + B_k u) \cdot y_{k,N} + ((K \star \mu_N)(x_i) + g(y)(x_{i,N})) \cdot x_{i,N} \\
&\leq \|(K \star \mu_N)(y_{k,N})\| \, \|y_{k,N}\| + f_k(y_N) \cdot y_{k,N} + \|B_k u\| \|y_{k,N}\| \\
&\quad + \|(K \star \mu_N)(x_{i,N})\| \|x_{i,N}\| + g(y_N)(x_{i,N}) \cdot x_{i,N} \\
&\leq C_K\left(1 + \|y_{k,N}\| + \frac{1}{N}\sum_{j=1}^N \|x_{j,N}\|\right)\|y_{k,N}\| + F_1 \max_{l=1,\ldots m}\|y_{l,N}\|^2 + F_2 + M_1\|y_{k,N}\| \\
&\quad + C_K\left(1 + \|x_{i,N}\| + \frac{1}{N}\sum_{j=1}^N \|x_{j,N}\|\right)\|x_{i,N}\| + G_1\|x_{i,N}\|^2 + G_2 \max_{l=1,\ldots m}\|y_{l,N}\|^2 + G_3 \\
&\leq C_1 \max_{(\ell,j)\in\Sigma}\left\{\|y_{\ell,N}\|^2 + \|x_{j,N}\|^2\right\} + C_2,
\end{aligned}
$$

with $C_1 = 4C_K + F_1 + G_2 + M_1$ and $C_2 = C_K + F_2 + G_3 + M_1$. If we denote with $b_{(k,i)}(t) = \|y_{k,N}(t)\|^2 + \|x_{i,N}(t)\|^2$ and with $a(t) = \max_{(l,j)\in\Sigma}\{b_{(l,j)}(t)\}$, then the Lipschitz continuity of a implies that a is a.e. differentiable, while by Stampacchia's Lemma (see for instance [28, Chapter 2, Lemma A.4]) for a.e. $t \in [0, T]$ there exists a $(l, j) \in \Sigma$ such that

$$\dot{a}(t) = \frac{d}{dt}\left(\|y_{l,N}(t)\|^2 + \|x_{j,N}(t)\|^2\right) \leq 2C_1 a(t) + 2C_2.$$

Hence, Gronwall's Lemma and Definition 1.4–(i) imply that

$$a(t) \leq (a(0) + 2C_2 t)e^{2C_1 t} \leq (C_0 + 2C_2 t)e^{2C_1 t}, \tag{2.2}$$

11

for some uniform constant C_0 only depending on y^0 and $\text{supp}(\mu^0)$. It then follows that the trajectories $(y_N(\cdot), \mu_N(\cdot))$ are bounded uniformly in N in a ball $B(0, \rho_T) \subset \mathbb{R}^d$, for

$$\rho_T := \sqrt{C_0 + 2C_2 T} e^{C_1 T},$$

that is positive and does not depend on t or on N. This in turn implies that the trajectories $(y_N(\cdot), \mu_N(\cdot))$ are uniformly Lipschitz continuous in N, as can be easily verified by computing $\|\dot{y}_{k,N}\|$ and $\|\dot{x}_{i,N}\|$ and noticing that all the functions involved are bounded by Hypotheses (H) and the fact that we are inside $B(0, \rho_T)$. Therefore

$$\|\dot{y}_{k,N}(t)\| \leq \rho'_T, \quad \|\dot{x}_{i,N}(t)\| \leq \rho'_T, \tag{2.3}$$

where the constant ρ'_T does not depend on t or on N.

By an application of the Ascoli-Arzelà theorem for functions on $[0, T]$ and values in the complete metric space $\mathcal{X}$, there exists a subsequence, again denoted by $(y_N(\cdot), \mu_N(\cdot))$ converging uniformly to a limit $(y(\cdot), \mu(\cdot))$, whose trajectories are also contained in $B(0, \rho_T)$. Due to the equi-Lipschitz continuity of $(y_N(\cdot), \mu_N(\cdot))$ and the continuity of the Wasserstein distance, we thus obtain for some $L_T > 0$

$$\|(y(t_2), \mu(t_2)) - (y(t_1), \mu(t_2))\|_{\mathcal{X}} = \lim_{N \to +\infty} \|(y_N(t_2), \mu_N(t_2)) - (y_N(t_1), \mu_N(t_1))\|_{\mathcal{X}} \leq L_T |t_2 - t_1|,$$

for all $t_1, t_2 \in [0, T]$. Hence, the limit trajectory $(y^*(\cdot), \mu^*(\cdot))$ belongs as well to $\text{Lip}([0, T]; \mathcal{X})$.

It is now necessary to show that the limit $(y(\cdot), \mu(\cdot))$ is a solution of (1.2). We first verify that y is a solution of the ODEs part for $\mu = \mu$. To this end, we observe that the limit $(y_N(\cdot), \mu_N(\cdot)) \to (y(\cdot), \mu(\cdot))$ in $\mathcal{X}$ specifies into

$$\begin{cases} y_N \rightrightarrows y, & \text{in } [0, T], \\ \dot{y}_N \rightharpoonup \dot{y}, & \text{in } L^1([0, T], \mathbb{R}^{2d}). \end{cases} \tag{2.4}$$

and

$$\lim_{N \to +\infty} \mathcal{W}_1(\mu_N(t), \mu(t)) = 0, \tag{2.5}$$

uniformly with respect to $t \in [0, T]$. As a consequence of (2.4), (2.5), hypothesis (K), and Lemma 2.5, for all $k = 1, \ldots, m$ we have in $[0, T]$ for $N \to +\infty$

$$\begin{array}{ccc} (K \star \mu_N)(y_{k,N}) & \rightrightarrows & (K \star \mu)(y_k), \\ f_k(y_N) & \rightrightarrows & f_k(y). \end{array} \tag{2.6}$$

To prove that $y(t)$ is actually the Carathéodory solution of (1.8), we have only to show that for all $k = 1, \ldots, m$ one has

$$\dot{y}_k = (K \star \mu)(y_k) + f_k(y) + B_k u.$$

This is clearly equivalent to the following: for every $\eta \in \mathbb{R}^d$ and every $\hat{t} \in [0, T]$ it holds

$$\eta \cdot \int_0^{\hat{t}} \dot{y}_k(t) \, dt = \eta \cdot \int_0^{\hat{t}} [(K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t))] \, dt, \tag{2.7}$$

which follows from (2.6) and from the weak L^1-convergence of $\dot{y}_{k,N}$ to $\dot{y}_k$ and of u_N to u for $N \to +\infty$.

We are now left with verifying that μ is a solution of (1.2) for $y = y$. For all $\hat{t} \in [0, T]$ and for all $\phi \in C_c^1(\mathbb{R}^d; \mathbb{R})$ we infer that

$$\langle \phi, \mu_N(\hat{t}) - \mu_N(0) \rangle = \int_0^{\hat{t}} \left[\int_{\mathbb{R}^d} \nabla \phi(x) \cdot [(K \star \mu_N)(x) + g(y_N)(x)] d\mu_N(t)(x) \right] dt,$$

12

which is verified by considering the differentiation

$$\frac{d}{dt}\langle\phi,\mu_N(t)\rangle \;=\; \frac{1}{N}\frac{d}{dt}\sum_{i=1}^{N}\phi(x_i(t)) = \frac{1}{N}\left[\sum_{i=1}^{N}\nabla\phi(x_i(t))\cdot\dot{x}_i(t)\right],$$

and directly applying the substitution $\dot{x}_i = (K\star\mu_N)(x_i) + g(y_N)(x_i)$. By Lemma 2.5 and (2.5), we also have that for every $\rho > 0$

$$\lim_{N\to+\infty}\|K\star\mu_N(t) - K\star\mu(t)\|_{L^\infty(B(0,\rho))} = 0 \text{ in } [0,T],$$

and, as $\phi\in C^1_c(\mathbb{R}^{2d})$ has compact support, it follows that

$$\lim_{N\to+\infty}\|\nabla\phi\cdot(K\star\mu_N(t) - K\star\mu(t))\|_\infty = 0 \text{ in } [0,T].$$

Similarly, we have

$$\lim_{N\to+\infty}\|\nabla\phi\cdot(g(y_N(t)) - g(y(t)))\|_\infty = 0 \text{ in } [0,T],$$

by the compact support of ϕ, the C^1-continuity of g and the uniform convergence of y_N to y. Denote with $\mathcal{L}^1{\llcorner}_{[0,\hat{t}]}$ the Lebesgue measure on the time interval $[0,\hat{t}]$. Since the product measures $\mathcal{L}^1{\llcorner}_{[0,\hat{t}]}\times\frac{1}{t}\mu_N(t)$ converge in $\mathcal{P}_1([0,\hat{t}]\times\mathbb{R}^{2d})$ to $\mathcal{L}^1{\llcorner}_{[0,\hat{t}]}\times\frac{1}{t}\mu(t)$, we finally get

$$\lim_{N\to+\infty}\int_0^{\hat{t}}\int_{\mathbb{R}^d}\nabla\phi(x)\cdot[K\star\mu_N(t)+g(y_N(t)](x)d\mu_N(t)(x)dt$$

$$= \int_0^{\hat{t}}\int_{\mathbb{R}^d}\nabla\phi(x)\cdot[K\star\mu(t)+g(y(t))](x)d\mu(t)(x)dt,$$

that, together with (2.7), proves that (y,μ) is a solution of (1.8) with initial data (y^0,μ^0) and control u. $\qquad\square$

Corollary 2.7. *Let $y^0\in\mathbb{R}^{dm}$, $\mu^0\in\mathcal{P}_c(\mathbb{R}^d)$, and $u\in L^1([0,T];\mathcal{U})$. Then, there exists a solution of (1.2) with control u and initial datum (y^0,μ^0).*

Proof. Follows from Proposition 2.6 by taking any sequence of empirical measures μ^0_N as in Definition 1.4–(i), and the constant sequence $u_N\equiv u$ for all $N\in\mathbb{N}$. $\qquad\square$

The following intermediate results shall be helpful in proving the continuous dependance on the initial data.

Proposition 2.8. *Let $K:\mathbb{R}^d\to\mathbb{R}^d$ and $g:\mathbb{R}^{dm}\to\mathcal{C}^2(\mathbb{R}^d;\mathbb{R}^d)$ satisfy hypotheses (K) and (G). Then, for every $R>0$, there exists $L'_R>0$ satisfying $L'_R\leq C'(1+R)$ for some $C'>0$, and*

$$\|(K\star\mu^1)(x^1) - (K\star\mu^2)(x^2)\| \leq L'_R(\mathcal{W}_1(\mu^1,\mu^2)+\|x^1-x^2\|), \tag{2.8}$$

for all $x^1,x^2\in B(0,R)\subset\mathbb{R}^d$ and $\mu^1,\mu^2\in\mathcal{P}_1(\mathbb{R}^d)$ with $\mathrm{supp}(\mu^1),\mathrm{supp}(\mu^2)\subseteq B(0,R)$.

Moreover, for every $R>0$, there exists $L_R>0$ satisfying $L_R\leq C(1+R)$ for some $C>0$, and

$$\|(K\star\mu^1)(x^1) + g(\overline{y})(x^1) - (K\star\mu^2)(x^2) - g(\overline{y})(x^2)\| \leq L_R(\mathcal{W}_1(\mu^1,\mu^2)+\|x^1-x^2\|), \tag{2.9}$$

for all $x^1,x^2\in B(0,R)\subset\mathbb{R}^d$, $\overline{y}\in B(0,R)\subset\mathbb{R}^{dm}$ and $\mu^1,\mu^2\in\mathcal{P}_1(\mathbb{R}^d)$ with $\mathrm{supp}(\mu^1),\mathrm{supp}(\mu^2)\subseteq B(0,R)$.

Proof. By hypothesis, we have

$$\|(K \star \mu^1)(x) - (K \star \mu^2)(x)\| = \left\| \int_{\mathbb{R}^n} K(x - x') \, d(\mu^1 - \mu^2)(x') \right\| \leq \mathrm{Lip}_{2R}(K) \mathcal{W}_1(\mu^1, \mu^2),$$

and

$$\|(K \star \mu^1)(x^1) - (K \star \mu^1)(x^2)\| \leq \int_{\mathbb{R}^n} \|K(x^1 - x) - K(x^2 - x)\| \, d\mu^1(x) \leq \mathrm{Lip}_{2R}(K) \|x^1 - x^2\|,$$

where $\mathrm{Lip}_{2R}(K)$ stands for the Lipschitz constant of K on $B(0, 2R)$. Since from (K) it follows

$$\mathrm{Lip}_{2R}(K) \leq 2C_K(1 + R),$$

this proves (2.8) for $L'_R := \mathrm{Lip}_{2R}(K)$ and $C' := C_K$. Moreover, there exists $\xi \in \{tx^1 + (1-t)x^2 : t \in [0,1]\}$ such that

$$\|g(\overline{y})(x^1) - g(\overline{y})(x^2)\| \leq \sup_{\xi \in B(0,R) \subset \mathbb{R}^d, \varsigma \in B(0,R) \subset \mathbb{R}^{dm}} \|\mathbf{D}_y g(\varsigma)(\xi)\| \|x^1 - x^2\| \leq M \|x^1 - x^2\|,$$

for some $M > 0$, from the regularity of g. It then suffices to observe that, for some $C > 0$, it holds

$$\mathrm{Lip}_{2R}(K) + M \leq 2C_K(1 + R) + M \leq C(1 + R).$$

This proves (2.9) for $L_R = \mathrm{Lip}_{2R}(K) + M$. □

The estimate in Proposition 2.8 shows that the following general result holds for vector fields of the form $v(t, x, \mu(t)) := (K \star \mu(t) + g(y(t)))(x)$, since from Proposition 2.6 follows that x, y and μ lie in domains with *a priori* known bounds.

Proposition 2.9. *Let $v, w : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}^d$ be vector fields that satisfy the following hypotheses:*

1. *v and w are measurable with respect to t;*

2. *for every $R > 0$ there exists L_R satisfying $L_R \leq C(1 + R)$ such that for all $\mu^1, \mu^2 \in \mathcal{P}_1(\mathbb{R}^d)$ with support in $B(0, R)$ and all $x^1, x^2 \in \mathbb{R}^d$ it holds*

$$\begin{aligned} \|v(t, x^1, \mu^1) - v(t, x^2, \mu^2)\| &\leq L_R(\mathcal{W}_1(\mu^1, \mu^2) + \|x^1 - x^2\|), \\ \|w(t, x^1, \mu^1) - w(t, x^2, \mu^2)\| &\leq L_R(\mathcal{W}_1(\mu^1, \mu^2) + \|x^1 - x^2\|). \end{aligned} \tag{2.10}$$

Moreover, given $\mu^{0,1}, \mu^{0,2} \in \mathcal{P}_c(\mathbb{R}^d)$, assume that there exist two corresponding solutions μ^1, μ^2 of (2.1) with vector fields v, w, respectively, and final time T. Then there exist constants C_1 and C_2 such that

$$\mathcal{W}_1(\mu^1(t), \mu^2(t)) \leq e^{C_1 t} \mathcal{W}_1(\mu^{0,1}, \mu^{0,2}) + \int_0^t C_2 e^{C_1 s} \sup_{x \in B(0,R)} \|v(s, x, \mu^1(s)) - w(s, x, \mu^2(s))\| \, ds, \tag{2.11}$$

where C_1 and C_2 depend on the final time T, on the radius R and L_R the Lipschitz constant in (2.10).

Proof. See proofs in [22, Lemma 6.5, Lemma 6.6, Theorem 6.8]. □

We now prove the continuous dependence on the initial data, that also gives uniqueness of the solution for (1.2).

Proposition 2.10. *Let the Hypotheses (H) hold. Let $u \in L^1([0,T], \mathcal{U})$ be given, and take two solutions (y^1, μ^1) and (y^2, μ^2) of (1.2) with control u and with initial data $(y^{0,1}, \mu^{0,1}), (y^{0,2}, \mu^{0,2}) \in \mathcal{X}$, respectively, where $\mu^{0,1}$ and $\mu^{0,2}$ have both compact support. Then there exists a constant $C_T > 0$ such that*

$$\|(y^1(t), \mu^1(t)) - (y^2(t), \mu^2(t))\|_{\mathcal{X}} \leq C_T \|(y^{0,1}, \mu^{0,1}) - (y^{0,2}, \mu^{0,2})\|_{\mathcal{X}}, \quad \text{for all } t \in [0,T]$$

Proof. We start by noticing that, by the definition of a solution, we infer the existence of a $\rho_T > 0$ for which $y^1(\cdot), y^2(\cdot) \in B(0, \rho_T) \subset \mathbb{R}^{dm}$ and $\operatorname{supp}(\mu^1(\cdot)), \operatorname{supp}(\mu^2(\cdot)) \subseteq B(0, \rho_T) \subset \mathbb{R}^d$.

We shall show the continuous dependence estimate by chaining the stability of the ODE

$$\dot{y}_k(t) = (K \star \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \dots, m, \tag{2.12}$$

with the one of the PDE

$$\partial_t \mu(t) = -\nabla_x \cdot [(K \star \mu(t) + g(y(t)))\mu(t)], \tag{2.13}$$

first addressing the dependence of (2.12). By integration we have

$$\|y_k^1(t) - y_k^2(t)\| \leq \|y_k^{0,1} - y_k^{0,2}\|$$
$$+ \int_0^t \left(\|(K \star \mu^1(s))(y_k^1(s)) - (K \star \mu^2(s))(y_k^2(s))\| + \|f_k(y^1(s)) - f_k(y^2(s))\| \right) ds. \tag{2.14}$$

For the sake of notation, we shall denote with

$$F = \max_{k=1,\dots,m} \operatorname{Lip}_{\rho_T}(f_k),$$
$$G = \sup_{\xi \in B(0,\rho_T) \subset \mathbb{R}^d, \varsigma \in B(0,\rho_T) \subset \mathbb{R}^{dm}} \|\mathbf{D}_y g(\varsigma)(\xi)\|.$$

For the left-hand side of (2.14), (2.8), the $\mathcal{C}^2$-regularity of f_k for $k = 1, \dots, m$, and the uniform bound on $y^1(\cdot)$ and $y^2(\cdot)$ yield

$$\|y_k^1(t) - y_k^2(t)\| \leq \|y_k^{0,1} - y_k^{0,2}\| + \tag{2.15}$$
$$+ \int_0^t \left(L'_{\rho_T} \mathcal{W}_1(\mu^1(s), \mu^2(s)) + L'_{\rho_T} \|y_k^1(s) - y_k^2(s)\| + F\|y^1(s) - y^2(s)\| \right) ds$$

We now consider (2.13). Define the vector fields

$$v^1(t, x, \mu) := (K \star \mu + g(y^1(t)))(x), \qquad v^2(t, x, \mu) := (K \star \mu + g(y^2(t)))(x),$$

and let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^\infty$ cutoff function on $B(0, \rho_T)$ with $\|\nabla \Phi\| \leq 1$ and compact support in $\mathbb{R}^d$. Observe that, since $\|y^1(\cdot)\|, \|y^2(\cdot)\| \leq \rho_T$ and $\operatorname{supp}(\mu^1(\cdot)), \operatorname{supp}(\mu^2(\cdot)) \subseteq B(0, \rho_T)$, then μ^1 and μ^2 also solve (2.1) with Φv^1 and Φv^2 in place of v^1 and v^2, respectively. It then follows easily from Proposition 2.8 that Proposition 2.9 holds for $v = \Phi v^1$ and $w = \Phi v^2$. Hence, from (2.11) and taking into account that $v^1 = \Phi v^1$ and $v^2 = \Phi v^2$ in $B(0, \rho_T)$, we have

$$\mathcal{W}_1(\mu^1(t), \mu^2(t)) \leq e^{C_1 t} \mathcal{W}_1(\mu^{0,1}, \mu^{0,2}) + \int_0^t C_2 e^{C_1 s} \sup_{x \in B(0, \rho_T)} \|v^1(s, x, \mu^1(s)) - v^2(s, x, \mu^2(s))\| \, ds,$$

for some constants C_1 and C_2. By (2.9) and the regularity of g, for every $s \in [0, T]$ we have

$$\|v^1(s, x, \mu^1(s)) - v^2(s, x, \mu^2(s))\| \leq L_{\rho_T} \mathcal{W}_1(\mu^1(s), \mu^2(s)) + G\|y^1(s) - y^2(s)\|.$$

15

This gives

$$\mathcal{W}_1(\mu^1(t), \mu^2(t)) \leq e^{C_1 t} \mathcal{W}_1(\mu^{0,1}, \mu^{0,2})$$
$$+ \int_0^t C_2 e^{C_1 s} \left(L_{\rho_T} \mathcal{W}_1(\mu^1(s), \mu^2(s)) + G\|y^1(s) - y^2(s)\| \right) ds. \qquad (2.16)$$

We now consider the function

$$\varepsilon(t) := \|(y^1(t), \mu^1(t)) - (y^2(t), \mu^2(t))\|_{\mathcal{X}}$$

and, combining (2.15) for each $k = 1, \ldots, m$ and (2.16), we obtain

$$\varepsilon(t) \leq \|y^{0,1} - y^{0,2}\| + \int_0^t \left(L'_{\rho_T} \mathcal{W}_1(\mu^1(s), \mu^2(s)) + L'_{\rho_T}\|y^1(s) - y^2(s))\| + mF\|y^1(s) - y^2(s)\| \right) ds$$

$$+ e^{C_1 t} \mathcal{W}_1(\mu^{0,1}, \mu^{0,2}) + \int_0^t C_2 e^{C_1 s} \left(L_{\rho_T} \mathcal{W}_1(\mu^1(s), \mu^2(s)) + G\|y^1(s) - y^2(s)\| \right) ds$$

$$\leq \varepsilon(0)e^{C_1 t} + \int_0^t (L'_{\rho_T} + mF + (L_{\rho_T} + G)C_2 e^{C_1 s})\varepsilon(s) \, ds.$$

Gronwall's lemma then implies

$$\varepsilon(t) \leq \varepsilon(0)e^{C_1 t} \left((L'_{\rho_T} + mF)t + \frac{(L_{\rho_T} + G)C_2}{C_1}(e^{C_1 t} - 1) \right).$$

Since $t \in [0, T]$, the result is proved. $\qquad \square$

Remark 2.11. Going back to the application of the Ascoli-Arzelá Theorem in Proposition 2.6, consider another converging subsequence of (y_N, μ_N). We can prove that its limit is another solution of (1.8). Since the solution is unique for Proposition 2.10, we have that all converging subsequences of (y_N, μ_N) have the same limit, hence the sequence (y_N, μ_N) has itself limit (y^*, μ^*).

Remark 2.12. Since equicompactly supported solutions are unique, given the initial datum, by Proposition 2.10, combined with Proposition 2.6 we infer that the support of the unique solution can be estimated as a function of the data, namely it is contained in a ball $B(0, \rho_T)$, where the constant is depending only on $y^0, \text{supp}(\mu^0), K, g, f_k, B_k, \mathcal{U}$, and T.

2.2 Existence and construction of solutions of Problem 1

In this section, we prove that Problem 1 admits a solution which is a mean-field optimal control. The proof generalizes similar results in [22].

We first recall the main definition of Γ-convergence. We then define the sequence of functionals $(F_N)_{N \in \mathbb{N}}$ related to Problem 2 and F related to Problem 1 and prove that $(F_N)_{N \in \mathbb{N}}$ Γ-converge to F.

Definition 2.13 (Γ-convergence). [20, Definition 4.1, Proposition 8.1] Let X be a metrizable separable space and $F_N \colon X \to (-\infty, \infty]$, $N \in \mathbb{N}$ be a sequence of functionals. We say that $(F_N)_{N \in \mathbb{N}}$ Γ-*converges* to F, written as $F_N \xrightarrow{\Gamma} F$, for a given $F \colon X \to (-\infty, \infty]$, if

1. $\liminf$-*condition:* For every $u \in X$ and every sequence $u_N \to u$,

$$F(u) \leq \liminf_{N \to +\infty} F_N(u_N);$$

2. $\limsup$-*condition:* For every $u \in X$, there exists a sequence $u_N \to u$, called *recovery sequence*, such that
$$F(u) \geq \limsup_{N \to +\infty} F_N(u_N).$$

Furthermore, we call the sequence $(F_N)_{N \in \mathbb{N}}$ *equi-coercive* if for every $c \in \mathbb{R}$ there is a compact set $K \subseteq X$ such that $\{u : F_N(u) \leq c\} \subseteq K$ for all $N \in \mathbb{N}$. As a direct consequence of equi-coercivity, assuming $u_N^* \in \arg\min F_N \neq \emptyset$ for all $N \in \mathbb{N}$, there is a subsequence $(u_{N_k}^*)_{k \in \mathbb{N}}$ and $u^* \in X$ such that
$$u_{N_k}^* \to u^* \in \arg\min F.$$

In view of the definition of Γ-convergence, let us fix as our domain $X = L^1([0,T];\mathcal{U})$ which, endowed with the weak L^1-topology, is actually a metrizable space.

Fix now an initial datum $(y^0, \mu^0) \in \mathcal{X}$, with μ^0 compactly supported, and choose a sequence μ_N^0 as in Definition 1.4–(i).

Consider the functional $F(u)$ on X defined in (1.1), where the pair (y, μ) defines the unique solution of (1.2) with initial datum (y^0, μ^0) and control u. Similarly, consider the functional $F_N(u)$ on X defined in (1.7), where the pair (y_N, μ_N) defines the unique solution of (1.2) with initial datum (y^0, μ_N^0) and control u. As recalled in Proposition 2.6, such solution coincides with the solution of the ODE system (1.8).

The rest of this section is devoted to the proof of the Γ-convergence of the sequence of functionals $(F_N)_{N \in \mathbb{N}}$ on X to the target functional F. Let us mention that Γ-convergence in optimal control problems has been already considered, see for instance [12], but, to our knowledge, it has been only recently specified in connection to mean-field limits in [22, 23].

Theorem 2.14. *Let the functionals* (1.1)-(1.7) *and dynamics* (1.2) *satisfy Hypotheses (H). Consider an initial datum* $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d)$, *and a sequence* $(\mu_N^0)_{N \in \mathbb{N}}$, *where* μ_N^0 *is as in Definition 1.4–(i). Then the sequence of functionals* $(F_N)_{N \in \mathbb{N}}$ *on* $X = L^1([0,T];\mathcal{U})$ *defined in* (1.7) Γ-*converges to the functional* F *defined in* (1.1).

Proof. Let us start by showing the $\Gamma - \liminf$ condition. Let us fix a weakly convergent sequence of controls $u_N \rightharpoonup u^*$ in X. We associate to each of these controls a sequence of solutions (y_N, μ_N) of (1.2) uniformly convergent to a solution (y^*, μ^*) with control u^* and initial datum (y^0, μ^0). In view of the fact that solutions (y_N, μ_N) and (y^*, μ^*) will have uniformly bounded supports with respect to N and $t \in [0,T]$ and by the uniform convergence of trajectories $y_N(t) \rightrightarrows y^*(t)$ as well as the uniform convergence $\mathcal{W}_1(\mu_N(t), \mu^*(t)) \to 0$ for $t \in [0,T]$, it follows from the continuity of L under Hypotheses (H) that
$$\lim_{N \to +\infty} \int_0^T L(y_N(t), \mu_N(t)) = \int_0^T L(y^*(t), \mu^*(t))dt. \tag{2.17}$$

By the assumed weak convergence of $(u_N)_{N \in \mathbb{N}}$ to $u^* \in X$ and Ioffe's Theorem (see, for instance, [2, Theorem 5.8]) we obtain the lower-semicontinuity of γ
$$\liminf_{N \to +\infty} \int_0^T \gamma(u_N(t))dt \geq \int_0^T \gamma(u^*(t))dt. \tag{2.18}$$

By combining (2.17) and (2.18), we immediately obtain the $\Gamma - \liminf$ condition
$$\liminf_{N \to +\infty} F_N(u_N) \geq F(u^*).$$

17

We now prove the $\Gamma - \limsup$ condition. We now fix u^* and consider the trivial recovery sequence $u_N \equiv u^*$ for all $N \in \mathbb{N}$. Similarly as above for the argument of the $\Gamma - \liminf$ condition, we can associate to each of these controls a sequence of solutions $(y_N(t), \mu_N(t))$ of (1.2) uniformly convergent to a solution $(y^*(t), \mu^*(t))$ with control u^* and initial datum (y^0, μ^0) and we can similarly conclude the limit (2.17). Additionally, since $(u_N)_{N \in \mathbb{N}}$ is a constant sequence, we have

$$\liminf_{N \to +\infty} \int_0^T \gamma(u_N(t)) dt = \int_0^T \gamma(u^*(t)) dt. \tag{2.19}$$

Hence, combining (2.17) and (2.19) we can easily infer

$$\limsup_{N \to \infty} F_N(u_N) = \lim_{N \to \infty} F_N(u^*) = F(u^*).$$

$\square$

Corollary 2.15. *Let the Hypotheses (H) in Section 1.1 hold. For every initial datum $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$, there exists a mean-field optimal control u^* for Problem 1.*

Proof. Consider empirical measures μ_N^0 as in Definition 1.4–(i). Notice that the optimal controls u_N^* of Problem 2 belong to $X = L^1([0,T];\mathcal{U})$, which is a compact set with respect to the weak topology of L^1. Hence, the sequence $(F_N)_{N \in \mathbb{N}}$ is equicoercive, and $(u_N^*)_{N \in \mathbb{N}}$ admits a subsequence, which we do not relabel, weakly convergent to some $u^* \in X$.

We can associate to each of these controls u_N^* and initial data (y^0, μ_N^0) a solution (y_N, μ_N) of (1.2). The sequence of solutions (y_N, μ_N) is then uniformly convergent to a solution (y^*, μ^*) of (1.2) with control u^*, by Proposition 2.6. In order to conclude that u^* is an optimal control for Problem 1 (and hence, by construction, that u^* is a *mean-field optimal control*) we need to show that it is actually a minimizer of F. For that we use the fact that F is the Γ-limit of the sequence $(F_N)_{N \in \mathbb{N}}$ as proved in Theorem 2.14. Let $u \in X$ be an arbitrary control and let $(u_N)_{N \in \mathbb{N}}$ be a recovery sequence given by the $\Gamma - \limsup$ condition, so that

$$F(u) \geq \limsup_{N \to +\infty} F_N(u_N). \tag{2.20}$$

By using now the optimality of $(u_N^*)_{N \in \mathbb{N}}$, we have

$$\limsup_{N \to +\infty} F_N(u_N) \geq \limsup_{N \to +\infty} F_N(u_N^*) \geq \liminf_{N \to +\infty} F_N(u_N^*). \tag{2.21}$$

Applying the $\Gamma - \liminf$ condition yields

$$\liminf_{N \to +\infty} F_N(u_N^*) \geq F(u^*). \tag{2.22}$$

By chaining the inequalities (2.20)-(2.21)-(2.22) we have

$$F(u) \geq F(u^*), \quad \text{for all } u \in X,$$

i.e., that u^* is an optimal control. $\square$

Remark 2.16. Observe that the previous result does not state uniqueness of the optimal control for the infinite dimensional problem. Indeed, in general, we cannot ensure that *all* solutions of Problem 1 are mean-field optimal controls.

3   The finite-dimensional problem

In this section we study the discrete Problem 2 and state the PMP for it. We first recall the following existence result for the optimal control problem.

Proposition 3.1 (Theorem 23.11, [15]). *Under Hypotheses (H), Problem 2 admits solutions.*

We now introduce the adjoint variables of x_i and y_k, denoted by p_i and q_k, respectively, and state the PMP in the following box.

Theorem 3.2 (Theorem 22.2, [15]). *Let u_N^* be a solution of Problem 2 with initial datum $(y(0), x(0)) = (y^0, x^0)$, and denote with $(y^*(\cdot), x^*(\cdot)) : [0, T] \to \mathbb{R}^{dm+dN}$ the corresponding trajectory. Then there exists a Lipschitz curve $(y^*(\cdot), q^*(\cdot), x^*(\cdot), p^*(\cdot)) \in \mathrm{Lip}([0, T], \mathbb{R}^{2dm+2dN})$ solving the system*

$$
\begin{cases}
\dot{y}_k^* = \nabla_{q_k} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) & \\
\dot{q}_k^* = -\nabla_{y_k} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) & k = 1, \ldots, m, \\
\dot{x}_i^* = \nabla_{p_i} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) & \\
\dot{p}_i^* = -\nabla_{x_i} \mathbb{H}_N(y^*, q^*, x^*, p^*, u^*) & i = 1, \ldots, N, \\
u_N^* = \arg\max_{u \in \mathcal{U}} \mathbb{H}_N(y^*, q^*, x^*, p^*, u),
\end{cases}
\tag{3.1}
$$

with initial datum $(y(0), x(0)) = (y^0, x^0)$ and terminal datum $(q(T), p(T)) = 0$, where the Hamiltonian $\mathbb{H}_N : \mathbb{R}^{2dm+2dN} \to \mathbb{R}$ is given by

$$
\begin{aligned}
\mathbb{H}_N(y, q, x, p, u) = &\sum_{i=1}^{N} p_i \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(x_i - x_j) + g(y)(x_i) \right) + \\
&+ \sum_{k=1}^{m} q_k \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(y_k - x_j) + f_k(y) + B_k u \right) - L(y, \mu_N) - \gamma(u),
\end{aligned}
\tag{3.2}
$$

with $\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i)$.

Remark 3.3. The general statement of the PMP contains both normal and abnormal minimizers. In our case, the simpler formulation of the PMP is given by the fact that we have normal minimizers only. This is a consequence of the fact that the final configuration is free, see e.g. [15, Corollary 22.3].

Remark 3.4. The uniqueness of the maximizer of $\mathbb{H}_N$ follows from the same motivations reported in Remark 1.2. Indeed, the form of the Hamiltonian implies that for each $u^* \in \mathcal{U}$ it holds

$$
u^* = \arg\max_{u \in \mathcal{U}} \mathbb{H}_N(y^*, q^*, x^*, p^*, u) \quad \text{when} \quad u^* = \arg\max_{u \in \mathcal{U}} \left(\sum_{k=1}^{m} q_k^* \cdot B_k u - \gamma(u) \right).
$$

In other terms, since the control acts on the y variables only, then we have a simpler formulation for the maximization of the Hamiltonian $\mathbb{H}_N$.

We now want to embed solutions of the PMP for Problem 2 as solutions of the extended PMP for Problem 1. As a first step, we prove that pairs control-trajectories $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying system (3.1) have support uniformly bounded in time and in $N \in \mathbb{N}$.

Proposition 3.5. *Let $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, and μ_N^0 be as in Definition 1.4–(i). Let u_N^* be a solution of Problem 2 with initial datum (y^0, μ_N^0), and let $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ be a pair control-trajectory satisfying the PMP for Problem 2 with initial datum (y^0, μ_N^0) and control u_N^* given by Theorem 3.2.*

Then the trajectories $(y_N^*(\cdot), q_N^*(\cdot), \nu_N^*(\cdot))$, where $\nu_N^* := \Pi_N(x_N^*, p_N^*)$, are equibounded and equi-Lipschitz continuous from $[0, T]$ to $\mathcal{Y}$, where the space $\mathcal{Y} := \mathbb{R}^{2dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ is endowed with the distance

$$\|(y, q, \nu) - (y', q', \nu')\|_{\mathcal{Y}} = \|y - y'\| + \|q - q'\| + \mathcal{W}_1(\nu, \nu'). \tag{3.3}$$

Furthermore, there exists $R_T > 0$, depending only on $y^0, \mathrm{supp}(\mu^0), d, K, g, f_k, B_k, \mathcal{U}$, and T, such that $\mathrm{supp}(\nu_N^*(\cdot)) \subseteq B(0, R_T)$ for all $N \in \mathbb{N}$. In particular, it holds $\mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*) = \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)$.

Proof. As a first step, notice that the pair (y_N^*, x_N^*) solves the system (1.8). It then follows from (2.2) and (2.3) that there exist two constants ρ_T and ρ_T', not depending on N such that, for all $i = 1, \dots, N$, for all $k = 1, \dots, m$, and a.e. $t \in [0, T]$ we have

$$\|y_{k,N}^*(t)\| \le \rho_T, \quad \|x_{i,N}^*(t)\| \le \rho_T \tag{3.4}$$

$$\|\dot{y}_{k,N}^*(t)\| \le \rho_T', \quad \|\dot{x}_{i,N}^*(t)\| \le \rho_T'. \tag{3.5}$$

It follows in particular that there exists a uniform constant W_T such that

$$\left\| \frac{1}{N} \sum_{i=1}^N \omega(x_{i,N}^*(t)) \right\| \le W_T$$

for all $t \in [0, T]$.

We now observe that by an explicit computation

$$
\begin{aligned}
N \nabla_{x_i} \mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \cdot e_l =& N \frac{r_{i,N}^*}{N} \cdot \frac{1}{N} \sum_{j=1}^N \mathbf{D}K(x_{i,N}^* - x_{j,N}^*) e_l - N \sum_{j=1}^N \frac{r_{j,N}^*}{N} \cdot \frac{1}{N} \mathbf{D}K(x_{j,N}^* - x_{i,N}^*) e_l \\
&+ N \frac{r_{i,N}^*}{N} \cdot \nabla g(y_N^*)(x_{i,N}^*) - N \sum_{k=1}^m q_{k,N}^* \cdot \frac{1}{N} \mathbf{D}K(y_{k,N}^* - x_{i,N}^*) e_l \\
&- N \frac{1}{N} \left(\nabla_\xi \ell \left(y_N^*, x_{i,N}^*, \tfrac{1}{N} \sum_{j=1}^N \omega(x_{j,N}^*) \right) \cdot e_l \right. \\
&+ \left. \nabla_\varsigma \ell \left(y_N^*, x_{i,N}^*, \tfrac{1}{N} \sum_{j=1}^N \omega(x_{j,N}^*) \right) \mathbf{D}\omega(x_{i,N}^*) \cdot e_l \right) \\
=& \frac{1}{N} \sum_{j=1}^N (r_{i,N}^* - r_{j,N}^*) \cdot \left(\mathbf{D}K(x_{i,N}^* - x_{j,N}^*) e_l \right) + r_{i,N}^* \cdot \mathbf{D}_x g(y_N^*)(x_{i,N}^*) \\
&- \sum_{k=1}^m q_{k,N}^* \cdot \left(\mathbf{D}K(y_{k,N}^* - x_{i,N}^*) e_l \right) - \nabla_\xi \ell \left(y_N^*, x_{i,N}^*, \tfrac{1}{N} \sum_{j=1}^N \omega(x_{j,N}^*) \right) \cdot e_l \\
&- \left(\nabla_\varsigma \ell \left(y_N^*, x_{i,N}^*, \tfrac{1}{N} \sum_{j=1}^N \omega(x_{j,N}^*) \right) \mathbf{D}\omega(x_{i,N}^*) \right) \cdot e_l,
\end{aligned}
$$

$$\tag{3.6}$$

for each $i = 1, \ldots, N$ and each $l = 1, \ldots, d$ (where we have used that $\mathbf{D}K$ is even and we merged the first two terms). Therefore, since $\dot{p}_{i,N}^*$ solves (3.1) and $\dot{r}_{i,N}^* = N\dot{p}_{i,N}^*$, we get

$$
\begin{aligned}
-\dot{r}_{i,N}^*(t) \cdot e_l = &\frac{1}{N} \sum_{j=1}^{N} (r_{i,N}^*(t) - r_{j,N}^*(t)) \cdot (\mathbf{D}K(x_{i,N}^*(t) - x_{j,N}^*(t))e_l) \\
&+ r_{i,N}^*(t) \cdot (\mathbf{D}_y g(y_N^*(t))(x_{i,N}^*(t))e_l) - \sum_{k=1}^{m} q_{k,N}^*(t) \cdot (\mathbf{D}K(y_{k,N}^*(t) - x_{i,N}^*(t))e_l) \\
&- \nabla_\xi \ell \left(y_N^*(t), x_{i,N}^*(t), \tfrac{1}{N} \sum_{j=1}^{N} \omega(x_{j,N}^*(t)) \right) \cdot e_l \\
&- \left(\nabla_\varsigma \ell \left(y_N^*(t), x_{i,N}^*(t), \tfrac{1}{N} \sum_{j=1}^{N} \omega(x_{j,N}^*(t)) \right) \mathbf{D}\omega(x_{i,N}^*(t)) \right) \cdot e_l,
\end{aligned}
$$

for each $i = 1, \ldots, N$, each $l = 1, \ldots, d$, and a.e. $t \in [0, T]$, where we have used the fact that $\mathbf{D}K$ is even. We now denote with L_T a uniform constant such that

$$
\|\mathbf{D}K\|_{L^\infty(B(0,\rho_T),\mathbb{R}^{d\times d})} \le L_T, \qquad \sup_{|y| \le \sqrt{m}R_T} \|\mathbf{D}_y g(y)(\cdot)\|_{L^\infty(B(0,\rho_T),\mathbb{R}^{d\times d})} \le L_T
$$

$$
\|\nabla_\xi \ell\|_{L^\infty(B(0,\sqrt{m}\rho_T) \times B(0,\rho_T) \times B(0,W_T),\mathbb{R}^d)} \le L_T, \qquad \|\nabla_\varsigma \ell\|_{L^\infty(B(0,\sqrt{m}\rho_T) \times B(0,\rho_T) \times B(0,W_T),\mathbb{R}^d)} \le L_T,
$$

$$
\|\mathbf{D}\omega\|_{L^\infty(B(0,\rho_T),\mathbb{R}^{d\times d})} \le L_T,
$$

and we easily get the estimate

$$
\|\dot{r}_{i,N}^*(t)\| \le \sqrt{d}L_T \left(2\|r_{i,N}^*(t)\| + \frac{1}{N} \sum_{j=1}^{N} \|r_{j,N}^*(t)\| + \sum_{k=1}^{m} \|q_{k,N}^*(t)\| + 1 + L_T \right) \tag{3.7}
$$

for each $i = 1, \ldots, N$ and a.e. $t \in [0, T]$. An explicit computation of $\nabla_{y_k} \mathbb{H}_N$ and a similar argument, possibly with another constant L_T, show the estimate

$$
\|\dot{q}_{k,N}^*(t)\| \le \sqrt{d}L_T \left(\frac{1}{N} \sum_{i=1}^{N} \|r_{i,N}^*(t)\| + 2\|q_{k,N}^*(t)\| + L_T \right) \tag{3.8}
$$

for each $k = 1, \ldots, m$ and a.e. $t \in [0, T]$. We now set

$$
\varepsilon_N(t) := \sum_{k=1}^{m} \|q_{k,N}^*(t)\| + \frac{1}{N} \sum_{i=1}^{N} \|r_{i,N}^*(t)\|,
$$

and observe that it holds

$$
|\dot{\varepsilon}_N(t)| \le \sum_{k=1}^{m} \|\dot{q}_{k,N}^*(t)\| + \frac{1}{N} \sum_{i=1}^{N} \|\dot{r}_{i,N}^*(t)\|,
$$

therefore (3.7) and (3.8) yield

$$
|\dot{\varepsilon}_N(t)| \le \sqrt{d}L_T \left(4\varepsilon_N(t) + 1 + 2L_T \right). \tag{3.9}
$$

Defining then the increasing functions $\eta_N(t)$ through $\eta_N(t) := \sup_{\tau \in [0,t]} \varepsilon_N(T - \tau)$, and observing that it holds $\eta_N(0) = 0$ for the boundary conditions in Theorem 3.2, from (3.9) and Gronwall's Lemma we obtain

$$
\eta_N(\tau) \le \sqrt{d}L_T\tau(1 + 2L_T)e^{(4\sqrt{d}L_T)\tau}
$$

21

and with this

$$\varepsilon_N(t) \le \eta_N(T) \le \sqrt{d} L_T T (1 + 2L_T) e^{(4\sqrt{d}L_T)T} := C_T \tag{3.10}$$

for all $t \in [0, T]$. Plugging into (3.9), we get the existence of a constant C_T' such that

$$|\dot{\varepsilon}_N(t)| \le C_T' \tag{3.11}$$

for a.e. $t \in [0, T]$. Since by definition of $\nu_N^*(t)$ and standard properties of the Wasserstein distance $\mathcal{W}_1$ it holds

$$\mathcal{W}_1(\nu_N^*(t + \tau), \nu_N^*(t)) \le \sqrt{2} \left(\frac{1}{N} \sum_{i=1}^{N} \|x_{i,N}^*(t + \tau) - x_{i,N}^*(t)\| + \frac{1}{N} \sum_{i=1}^{N} \|r_{i,N}^*(t + \tau) - r_{i,N}^*(t))\| \right),$$

from the previous inequality, (3.4), (3.5), (3.9), (3.10), and (3.11) we obtain that $y_N^*(t)$ and $q_N^*(t)$ are equibounded, that there exist a constant, denoted by R_T, such that $\text{supp}(\nu_N^*(t)) \subset B(0, R_T)$ for all $t \in [0, T]$ and that (y_N^*, q_N^*, ν_N^*) are equi-Lipschitz continuous from $[0, T]$ with values in $\mathcal{Y}$. $\square$

Proposition 3.6. *Let $N \in \mathbb{N}$ and $u_N^* \in L^p([0, T]; \mathcal{U})$ be an optimal control for Problem 2 with given initial datum $(y_N^0, x_N^0) \in \mathbb{R}^{dm + dN}$, and $(y_N^*(\cdot), q_N^*(\cdot), x_N^*(\cdot), p_N^*(\cdot)) \in \text{Lip}([0, T], \mathbb{R}^{2dm + 2dN})$ a corresponding trajectory of the PMP with maximized Hamiltonian $\mathbb{H}_N$.*

Define $\nu_N^* := \Pi_N(x_{1,N}^*, p_{1,N}^*, \ldots, x_{N,N}^*, p_{N,N}^*)$ with Π_N as in (1.11), and assume that $\text{supp}(\nu_N^*(\cdot)) \subseteq B(0, R_T)$. Then, the control u_N^* is optimal for Problem 1 and $(y_N^*, q_N^*, \nu_N^*, u_N^*)$ satisfies the extended Pontryagin Maximum Principle.

Proof. First observe that, by Proposition 3.5, $\mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*) = \mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*)$ and that for every $t \in [0, T]$

$$u_N^*(t) = \arg \max_{u \in \mathcal{U}} \mathbb{H}_N(y_N^*(t), q_N^*(t), x_N^*(t), p_N^*(t), u) \iff u_N^*(t) = \arg \max_{u \in \mathcal{U}} \mathbb{H}(y_N^*(t), q_N^*(t), \nu_N^*(t), u),$$

due to the specific form of the Hamiltonian $\mathbb{H}_N$ and $\mathbb{H}$, see Remark 3.4.

We now prove that

$$\begin{cases} \dot{y}_{k,N}^* &= \nabla_{q_k} \mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*), \\ \dot{q}_{k,N}^* &= -\nabla_{y_k} \mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*), \end{cases} \implies \begin{cases} \dot{y}_{k,N}^* &= \nabla_{q_k} \mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*), \\ \dot{q}_{k,N}^* &= -\nabla_{y_k} \mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*), \end{cases}$$

i.e., that if the (y, q) variables satisfy the PMP for Problem 2 then they satisfy the extended PMP for Problem 1. It is sufficient to observe that $\mathbb{H}_N$ can be rewritten in terms of $\nu_N^*(\cdot)$ as follows

$$\begin{aligned} \mathbb{H}_N(y_N^*, q_N^*, \nu_N^*, u_N^*) &= \int_{\mathbb{R}^{4d}} r \cdot (K \star \pi_{1\#} \nu_N^*)(x) \, d\nu_N^*(x, r) + \int_{\mathbb{R}^{2d}} r \cdot g(y_N^*)(x) \, d\nu_N^*(x, r) \\ &+ \sum_{k=1}^{m} \int_{\mathbb{R}^{2d}} q_{k,N}^* \cdot K(y_{k,N}^* - x) \, d\nu_N^*(x, r) + \sum_{k=1}^{m} q_{k,N}^* \cdot (f_k(y_{k,N}^*) + B_k u_N^*) \\ &- L(y_N^*, \pi_{1\#} \nu_N^*) - \gamma(u_N^*), \end{aligned}$$

where we used the variable $r = Np$. Comparing it with $\mathbb{H}(y_N^*, q_N^*, \nu_N^*, u_N^*)$, one has that their expressions coincide up to the first term. Since such first term is independent on y_k and q_k, then $\nabla_{y_k} \mathbb{H}_N = \nabla_{y_k} \mathbb{H}$ and $\nabla_{q_k} \mathbb{H}_N = \nabla_{q_k} \mathbb{H}$, hence equations for $\dot{y}_{k,N}^*, \dot{q}_{k,N}^*$ in the PMP for Problem 2 and in the extended PMP for Problem 1 coincide.

We now prove a similar result for the $(x_{i,N}^*, r_{i,N}^*)$ variables, with $r_{i,N}^* = Np_{i,N}^*$. After this change of variable, the third and the fourth equation in (3.1) become

$$\begin{cases} \dot{x}_{i,N}^* &= N\nabla_{r_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \\ \dot{r}_{i,N}^* &= -N\nabla_{x_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*). \end{cases}$$

We want to prove that the following identity holds

$$J(\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*))(x_{i,N}^*, r_{i,N}^*) = \begin{pmatrix} N\nabla_{r_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \\ -N\nabla_{x_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \end{pmatrix}, \tag{3.12}$$

i.e., that the Hamiltonian vector fields generated by $\mathbb{H}$ and $\mathbb{H}_N$ coincide in each point $(x_{i,N}^*, r_{i,N}^*)$. The presence of the constant N in the right-hand side is due to the change of variables $r_{i,N}^* = Np_{i,N}^*$. By applying J^{-1} on both sides of (3.12), we need to prove

$$\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)(x_{i,N}^*, r_{i,N}^*) \cdot e_l = N\nabla_{x_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \cdot e_l \qquad \text{for } l = 1, \ldots, d, \tag{3.13}$$
$$\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)(x_{i,N}^*, r_{i,N}^*) \cdot e_l = N\nabla_{r_i}\mathbb{H}_N(y_N^*, q_N^*, x_N^*, p_N^*, u_N^*) \cdot e_{l-d} \qquad \text{for } l = d+1, \ldots, 2d. \tag{3.14}$$

By writing explicitly the left hand sides of (3.13) and (3.14) by using the expressions (1.5)-(1.6) and evaluating them in $(x_{i,N}^*, r_{i,N}^*)$, we have

$$\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)(x_{i,N}^*, r_{i,N}^*) \cdot e_l = \frac{1}{N}\sum_{j=1}^N (r_{i,N}^* - r_{j,N}^*) \cdot \left(\mathbf{D}K(x_{i,N}^* - x_{j,N}^*)e_l\right) + r_{i,N}^* \cdot \left(\mathbf{D}_x g(y_N^*)(x_{i,N}^*)e_l\right)$$

$$- \sum_{k=1}^m q_{k,N}^* \cdot \left(\mathbf{D}K(y_{k,N}^* - x_{i,N}^*)e_l\right) - \nabla_\xi \ell\left(y_N^*, x_{i,N}^*, \frac{1}{N}\sum_{j=1}^N \omega(x_{j,N}^*)\right) \cdot e_l$$

$$- \left(\nabla_\varsigma \ell\left(y_N^*, x_{i,N}^*, \frac{1}{N}\sum_{j=1}^N \omega(x_{j,N}^*)\right)\mathbf{D}\omega(x_{i,N}^*)\right) \cdot e_l,$$

for $l = 1, \ldots, d$, so that (3.13) follows immediately from (3.6). Similarly, we have

$$\nabla_\nu \mathbb{H}_c(y_N^*, q_N^*, \nu_N^*, u_N^*)(x_{i,N}^*, r_{i,N}^*) \cdot e_l = \frac{1}{N}\sum_{j=1}^N K(x_{i,N}^* - x_{j,N}^*) \cdot e_{l-d} + g(y_N^*)(x_{i,N}^*) \cdot e_{l-d}$$

for $l = d+1, \ldots, 2d$, which coincides with the right hand side of (3.14) by an explicit computation. Since the boundary conditions of Problem 2 and Problem 1 coincide too, after the identification $\nu_N^* := \Pi_N(x_{1,N}^*, p_{1,N}^*, \ldots, x_{N,N}^*, p_{N,N}^*)$, the result follows now by (3.13)-(3.14) arguing, for instance, as in [23, Lemma 4.3]. $\square$

Remark 3.7. It is interesting to observe that the embedding of a trajectory of the PMP for Problem 2 to the empirical measure formulation depends on the number N of agents, see the definition of Π_N in (1.11). This is a consequence of the fact that the Hamiltonian of the PMP for Problem 2 actually depends on the number of agents. Indeed, consider a population composed of a unique agent (x_1, p_1), for which the first term of the Hamiltonian reads as $p_1 \cdot g(y)(x_1)$. Consider now a population composed of two agents (x_1, p_1, x_2, p_2) satisfying $x_1 = x_2$ and $p_1 = p_2$, for which the first term of the Hamiltonian reads as $2p_1 \cdot g(y)(x_1)$.

Clearly, in both cases the empirical measure in the state variables is $\mu_1 = \mu_2 = \delta(x - x_1)$, while the definition of Π_N gives two different empirical measures for the cotangent bundle: $\nu_1 = \delta(x - x_1, r - p_1)$ and $\nu_2 = \frac{1}{2}\delta(x - x_1, r - 2p_1)$. This difference is needed to compensate the dependence of the Hamiltonian of the PMP for Problem 2 on the number N of agents.

4   The Wasserstein gradient

We anticipated in Section 1 that the dynamics of ν^* in (1.3) is an Hamiltonian flow in the Wasserstein space of probability measures, in the sense of [3]. This means that the vector field $\nabla_\nu \mathbb{H}_c(\nu^*)$ is an element with minimal norm in the Fréchet subdifferential at the point ν^* of the maximized Hamiltonian $\mathbb{H}_c$ introduced in Theorem 1.1 (we drop for simplicity the y, q and u dependency). The proof of this fact shall follow the strategy adopted to obtain analogous results in [4, Chapter 10], which however cannot be applied verbatim to our case due to the peculiar nature of our operators. In order to use those techniques, we consider our functionals defined on $\mathcal{P}_2(\mathbb{R}^{2d})$ instead than on $\mathcal{P}_1(\mathbb{R}^{2d})$. Since we have already proven in Proposition 3.5 that, whenever we start from a compactly supported initial datum, the dynamics remains compactly supported uniformly in time, this assumption does not alter our conclusions.

We start with some basic definitions and general results on functionals defined on $\mathcal{P}_2(\mathbb{R}^{2d})$: the following one is motivated by Definition 10.3.1 and Remark 10.3.3 in [4].

Definition 4.1. Let $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to (-\infty, +\infty]$ be a proper and lower semicontinuous functional, and let $\nu_0 \in D(\psi)$. We say that $w \in L^2_{\nu_0}(\mathbb{R}^{2d})$ belongs to the *(Fréchet) subdifferential of ψ at ν_0*, in symbols $w \in \partial \psi(\nu_0)$ if and only if for any $\nu_1 \in \mathcal{P}_2(\mathbb{R}^{2d})$ it holds

$$\psi(\nu_1) - \psi(\nu_0) \geq \inf_{\rho \in \Gamma_o(\nu_0, \nu_1)} \int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) + o(\mathcal{W}_2(\nu_1, \nu_0)).$$

Proposition 4.2 ([4], Theorem 10.3.10)**.** *Fix the functional* $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to (-\infty, +\infty]$. *Then, for every* $\nu_0 \in D(\psi)$, *the metric slope*

$$|\partial \psi|(\nu_0) = \limsup_{\nu_1 \to \nu_0} \frac{(\psi(\nu_1) - \psi(\nu_0))^+}{\mathcal{W}_2(\nu_1, \nu_0)}$$

satisfies

$$|\partial \psi|(\nu_0) \leq \|w\|_{L^2_{\nu_0}} \tag{4.1}$$

for every $w \in \partial \psi(\nu_0)$.

The following property shall guarantee that the subdifferential of $\mathbb{H}_c$ is nonempty.

Definition 4.3. A proper, lower semicontinuous functional $\psi : \mathcal{P}_2(\mathbb{R}^n) \to (-\infty, +\infty]$ is *semiconvex along geodesics* whenever, for every $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ and $\rho \in \Gamma_o(\nu_0, \nu_1)$ there exists $C \in \mathbb{R}$ for which it holds

$$\psi(((1-s)\pi_1 + s\pi_2)_{\#}\rho) \leq (1-s)\psi(\nu_0) + s\psi(\nu_1) + Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1)) \text{ for every } s \in [0,1].$$

In what follows, we shall fix $y, q \in \mathbb{R}^{dm}$ and $u \in L^1([0,T]; \mathcal{U})$ and we write, for the sake of compactness, $\mathbb{H}_c(\nu)$ in place of $\mathbb{H}_c(y, q, \nu, u)$. Moreover, $\mathcal{K}$ shall denote a convex, compact subset of $\mathbb{R}^{2d}$ and $z = (x, r)$ a variable in $\mathbb{R}^{2d}$.

Whenever $\operatorname{supp}(\nu) \subseteq \overline{B(0, R_T)}$, $\mathbb{H}_c(\nu)$ can be rewritten as

$$\mathbb{H}_c(\nu) = \frac{1}{2} \int_{\mathbb{R}^{4d}} \mathcal{F}(z - z') d\nu(z)\nu(z') + \int_{\mathbb{R}^d} \mathcal{G}(z) d\nu(z) - \int_{\mathbb{R}^d} \ell(\pi_1(z), \int \omega \pi_{1\#}\nu) d\nu(z) + Q,$$

where

$$\mathcal{F}(x, r) = r \cdot K(x)$$

$$\mathcal{G}(x, r) = r \cdot g(y)(x) + \sum_{k=1}^m q_k \cdot K(y_k - x),$$

24

and Q collects all the remaining terms not depending on ν. Notice that $\mathcal{F}$ is an even function.

In order to prove the semiconvexity of $\mathbb{H}_c$, we shall establish the semiconvexity of the following functionals:

$$\hat{\mathbb{H}}_c^1(\nu) = \frac{1}{2} \int_{\mathbb{R}^{4d}} \hat{\mathcal{F}}(z - z')d\nu(z)\nu(z') + \int_{\mathbb{R}^d} \hat{\mathcal{G}}(z)d\nu(z),$$

$$\hat{\mathbb{H}}_c^2(\nu) = \int_{\mathbb{R}^d} \hat{\ell}(z, \int \hat{\omega}\nu)d\nu(z),$$

where $\hat{\mathcal{F}}$, $\hat{\mathcal{G}}$, $\hat{\ell}$, and $\hat{\omega}$ are $\mathcal{C}^2$ functions. The desired result will then follow by noticing that $\mathbb{H}_c(\nu) = \hat{\mathbb{H}}_c^1(\nu) + \hat{\mathbb{H}}_c^2(\nu)$ for $\hat{\mathcal{F}} = \mathcal{F}$, $\hat{\mathcal{G}} = \mathcal{G}$, $\hat{\ell} = -\ell \circ (\pi_1, \mathrm{Id})$, $\hat{\omega} = \omega \circ \pi_1$ and $\mathcal{K} = \overline{B(0, R_T)}$.

The following simple property will be needed to prove semiconvexity of the above functionals.

Lemma 4.4. Let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ with support contained in $\mathcal{K}$. Let $\rho \in \Gamma(\nu_0, \nu_1)$ and set

$$\nu_s = ((1 - s)\pi_1 + s\pi_2)_\# \rho, \tag{4.2}$$

for every $s \in [0, 1]$. Then, it holds

$$\mathrm{supp}(\nu_s) \subseteq \mathcal{K} \quad \text{for all } s \in [0, 1].$$

Proof. We first notice, that for every $\rho \in \Gamma(\nu_0, \nu_1)$ it holds

$$\mathrm{supp}(\rho) \subseteq \mathcal{K} \times \mathcal{K}. \tag{4.3}$$

This follows from the equality

$$\mathbb{R}^{4d} \backslash (\mathcal{K} \times \mathcal{K}) = (\mathbb{R}^{2d} \times (\mathbb{R}^{2d} \backslash \mathcal{K})) \cup ((\mathbb{R}^{2d} \backslash \mathcal{K}) \times \mathbb{R}^{2d})$$

and from the fact that both $\mathbb{R}^{2d} \times (\mathbb{R}^{2d} \backslash \mathcal{K}))$ and $(\mathbb{R}^{2d} \backslash \mathcal{K}) \times \mathbb{R}^{2d}$ are ρ-null sets by hypothesis.

To prove the Lemma, it suffices to show that for all $f \in \mathcal{C}(\mathbb{R}^{2d})$ satisfying $f \equiv 0$ on $\mathcal{K}$ it holds

$$\int_{\mathbb{R}^{2d}} f d\nu_s = 0. \tag{4.4}$$

Indeed,

$$\int_{\mathbb{R}^{2d}} f d\nu_s = \int_{\mathbb{R}^{4d}} f d((1 - s)\pi_1 + s\pi_2)_\# \rho(z_0, z_1)$$

$$= \int_{\mathbb{R}^{4d}} f((1 - s)z_0 + sz_1)d\rho(z_0, z_1)$$

$$= \int_{\mathcal{K} \times \mathcal{K}} f((1 - s)z_0 + sz_1)d\rho(z_0, z_1),$$

since, by (4.3), $\mathrm{supp}(\rho) \subseteq \mathcal{K} \times \mathcal{K}$. From the convexity of $\mathcal{K}$ follows that $(1 - s)z_0 + sz_1 \in \mathcal{K}$ for every $s \in [0, 1]$, which, together with the assumption $f \equiv 0$ in $\mathcal{K}$, yield (4.4), as desired. $\square$

In what follows, we shall make use of the following, well-known result.

Remark 4.5. Let $\mathcal{K}$ be a convex, compact subset of $\mathbb{R}^{2d}$ and let $f \in \mathcal{C}^2(\mathbb{R}^{2d}; \mathbb{R})$. Then there exists $C_{\mathcal{K}, f} \in \mathbb{R}$ depending only on $\mathcal{K}$ and f such that

$$f((1 - s)x_0 + sx_1) \leq (1 - s)f(x_0) + sf(x_1) + C_{\mathcal{K}, f}s(1 - s)\|x_0 - x_1\|^2, \tag{4.5}$$

for every $x_0, x_1 \in \mathbb{R}^{2d}$ and $s \in [0, 1]$.

We now prove the semiconvexity of $\hat{\mathbb{H}}^1_c$.

Lemma 4.6. *Let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ and let $\rho \in \Gamma(\nu_0, \nu_1)$. Then, there exists $C \in \mathbb{R}$ independent of ν_0 and ν_1 for which*

$$\hat{\mathbb{H}}^1_c(((1-s)\pi_1 + s\pi_2)_{\#}\rho) \leq (1-s)\hat{\mathbb{H}}^1_c(\nu_0) + s\hat{\mathbb{H}}^1_c(\nu_1) + Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1)$$

holds for every $s \in [0,1]$.

Proof. We may assume $\mathrm{supp}(\nu_0), \mathrm{supp}(\nu_1) \subseteq \mathcal{K}$ for some convex and compact set $\mathcal{K} \subset \mathbb{R}^{2d}$, otherwise the inequality is trivial. Hence, from Lemma 4.4, it follows $\mathrm{supp}(\nu_s) \subseteq \mathcal{K}$ for every $s \in [0,1]$. But then, since $\hat{\mathcal{F}}$ and $\hat{\mathcal{G}}$ are both $\mathcal{C}^2$, the result follows as in [4, Proposition 9.3.2, Proposition 9.3.5]. $\square$

Corollary 4.7. *Let $\hat{\omega} \in \mathcal{C}^2(\mathbb{R}^{2d}; \mathbb{R}^d)$, $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Gamma(\nu_0, \nu_1)$ and define ν_s as in (4.2) for $s \in [0,1]$. If we set*

$$\xi_s = \int_{\mathbb{R}^{2d}} \hat{\omega} d\nu_s, \tag{4.6}$$

then

$$\|\xi_s - (1-s)\xi_0 - s\xi_1\| \leq Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1),$$

for all $s \in [0,1]$, where C is independent of ν_0 and ν_1.

Proof. Follows from Lemma 4.6 applied first to the functions $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} \equiv \hat{\omega}$, and then to $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} \equiv -\hat{\omega}$. $\square$

The semiconvexity of $\hat{\mathbb{H}}^2_c$ will be deduced as a corollary of the following estimate.

Lemma 4.8. *Suppose that $\hat{\ell} \in \mathcal{C}^2(\mathbb{R}^{2d} \times \mathbb{R}^d; \mathbb{R})$, let $z_0, z_1 \in \mathcal{K}$ and set $z_s = (1-s)z_0 + sz_1$ for all $s \in [0,1]$. Furthermore, let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Gamma(\nu_0, \nu_1)$ and define ν_s and ξ_s as in (4.2) and (4.6) for $s \in [0,1]$. Then, for all $s \in [0,1]$, it holds*

$$\hat{\ell}(z_s, \xi_s) \leq (1-s)\hat{\ell}(z_0, \xi_0) + s\hat{\ell}(z_1, \xi_1) + C_{\mathcal{K}, \hat{\ell}, \hat{\omega}} s(1-s)\mathcal{W}_2^2(\nu_0, \nu_1) + C_{\mathcal{K}, \hat{\ell}, \hat{\omega}} s(1-s)\|z_0 - z_1\|^2,$$

for some constant $C_{\mathcal{K}, \hat{\ell}, \hat{\omega}}$ depending only on $\mathcal{K}, \hat{\ell}$ and $\hat{\omega}$.

Proof. Since $\mathcal{K}$ is compact, $z_s \in \mathcal{K}$ for all $s \in [0,1]$. Moreover, $(1-s)\xi_0 + s\xi_1 \in \mathcal{K}'$ for all $s \in [0,1]$, for some convex and compact set $\mathcal{K}' \subset \mathbb{R}^d$. Notice that from (4.5) follows

$$\hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1) \leq (1-s)\hat{\ell}(z_0, \xi_0) + s\hat{\ell}(z_1, \xi_1) + C_{\mathcal{K}, \mathcal{K}'} s(1-s)\left(\|z_0 - z_1\|^2 + \|\xi_0 - \xi_1\|^2\right), \quad (4.7)$$

and from the definition of ξ_s and Jensen's inequality, we get

$$\|\xi_0 - \xi_1\|^2 \leq \mathrm{Lip}_{\mathcal{K}}(\omega)\mathcal{W}_1^2(\nu_0, \nu_1) \leq \mathrm{Lip}_{\mathcal{K}}(\omega)\mathcal{W}_2^2(\nu_0, \nu_1). \tag{4.8}$$

Moreover, for every $s \in [0,1]$ it holds

$$\begin{aligned}
\|\hat{\ell}(z_s, \xi_s) - \hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1)\| &\leq \mathrm{Lip}_{\mathcal{K} \times \mathcal{K}'}\|\xi_s - (1-s)\xi_0 - s\xi_1\| \\
&\leq \mathrm{Lip}_{\mathcal{K} \times \mathcal{K}'} s(1-s)C\mathcal{W}_2^2(\nu_0, \nu_1).
\end{aligned} \tag{4.9}$$

Hence, for every $s \in [0,1]$, using (4.7), (4.8) and (4.9), we get

$$\begin{aligned}
\hat{\ell}(z_s, \xi_s) &= \hat{\ell}(z_s, \xi_s) - \hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1) + \hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1) \\
&\leq (1-s)\hat{\ell}(z_0, \xi_0) + s\hat{\ell}(z_1, \xi_1) + C_{\mathcal{K}, \hat{\ell}, \hat{\omega}} s(1-s)\mathcal{W}_2^2(\nu_0, \nu_1) + C_{\mathcal{K}, \hat{\ell}, \hat{\omega}} s(1-s)\|z_0 - z_1\|^2.
\end{aligned}$$

$\square$

Corollary 4.9. *Let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ and $\rho \in \Gamma_o(\nu_0, \nu_1)$. Then, there exists $C \in \mathbb{R}$ independent of ν_0 and ν_1 for which*

$$\hat{\mathbb{H}}_c^2(((1-s)\pi_1 + s\pi_2)_\#\rho) \leq (1-s)\hat{\mathbb{H}}_c^2(\nu_0) + s\hat{\mathbb{H}}_c^2(\nu_1) + Cs(1-s)\mathcal{W}_2^2(\nu_0, \nu_1)$$

holds for every $s \in [0,1]$.

Proof. Notice that, by Lemma 4.4, $\hat{H}_c^2(\nu_s)$ can be rewritten as

$$\hat{\mathbb{H}}_c^2(\nu_s) = \int_{\mathcal{K} \times \mathcal{K}} \hat{\ell}(z_s, \xi_s) d\rho(z_0, z_1),$$

Furthermore, since $\rho \in \Gamma_o(\nu_0, \nu_1)$ it holds

$$\int_{\mathcal{K} \times \mathcal{K}} \|z_0 - z_1\|^2 d\rho(z_0, z_1) = \int_{\mathbb{R}^{4d}} \|z_0 - z_1\|^2 d\rho(z_0, z_1) = \mathcal{W}_2^2(\nu_0, \nu_1),$$

the thesis follows from Lemma 4.8. $\qquad\square$

Proposition 4.10. *The functional $\mathbb{H}_c$ is semiconvex along geodesics.*

Proof. Follows directly from Lemma 4.6 and Corollary 4.9, by noticing that $\mathbb{H}_c(\nu) = \hat{\mathbb{H}}_c^1(\nu) + \hat{\mathbb{H}}_c^2(\nu)$ for $\hat{\mathcal{F}} = \mathcal{F}$, $\hat{\mathcal{G}} = \mathcal{G}$, $\hat{\ell} = -\ell \circ (\pi_1, \mathrm{Id})$, $\hat{\omega} = \omega \circ \pi_1$ and $\mathcal{K} = \overline{B(0, R_T)}$. $\qquad\square$

We define the vector field $\nabla_\nu \mathcal{L} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ as

$$\nabla_\nu \mathcal{L}(z) = \begin{bmatrix} \nabla_\xi \ell(y, \pi_1(z), \int \omega \pi_{1\#}\nu) + \nabla_\varsigma \ell(y, \pi_1(z), \int \omega \pi_{1\#}\nu)\mathbf{D}\omega(\pi_1(z)) \\ 0 \end{bmatrix},$$

for every $z \in \mathbb{R}^{2d}$. This notation is reminiscent of the fact that this vector field will eventually turn out to be the 2-Wasserstein gradient of the functional L, as it will follow from Theorem 4.12 in the case $\mathcal{F} \equiv \mathcal{G} \equiv 0$. We can thus define our candidate vector field for the Wasserstein gradient $\nabla_\nu \mathbb{H}_c(\nu_0)$ in the case that $\mathrm{supp}(\nu_0) \subseteq B(0, R_T)$:

$$w := (\nabla\mathcal{F}) \star \nu + \nabla\mathcal{G} - \nabla_\nu\mathcal{L}. \tag{4.10}$$

Notice that, by Hypotheses (H), w is a continuous function in z, and hence it is well-defined ν-a.e..

Lemma 4.11. *Let $\nu \in \mathcal{P}_c(\mathbb{R}^{2d})$. Then w defined by (4.10) belongs to $L_\nu^p(\mathbb{R}^{2d})$ for every $p \in [1, +\infty]$, and it satisfies*

$$\int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) = \int_{\mathbb{R}^{6d}} (\nabla\mathcal{F}(z_0 - z_2) + \nabla\mathcal{G}(z_0) - \nabla_\nu\mathcal{L}(z_0)) \cdot (z_1 - z_0) d\rho(z_0, z_1) d\nu(z_2) \tag{4.11}$$

for every plan $\rho \in \Gamma(\nu, \nu')$ such that $\nu' \in \mathcal{P}_c(\mathbb{R}^{2d})$.

Proof. Since w is continuous, the fact that w is L_ν^p-integrable follows the fact that ν has compact support. Equation (4.11) then follows by Fubini-Tonelli and from the fact that ρ is compactly supported too by Remark 4.4. $\qquad\square$

Theorem 4.12. *Let $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ be such that $\mathrm{supp}(\nu) \subseteq B(0, R_T)$. Then $\nu \in D(|\partial\mathbb{H}_c|)$ if and only if w as in (4.10) belongs to $L_\nu^2(\mathbb{R}^{2d})$. In this case, $\|w\|_{L_\nu^2} = |\partial\mathbb{H}_c|(\nu)$, i.e., w is an element with minimal norm in $\partial\mathbb{H}_c(\nu)$.*

Proof. We start by assuming that $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ satisfies $|\partial \mathbb{H}_c|(\nu) < +\infty$ and proving that this implies that w belongs to $L^2_\nu(\mathbb{R}^{2d})$ and that $\|w\|_{L^2_\nu} \le |\partial \mathbb{H}_c|(\nu)$. We compute the directional derivative of $\mathbb{H}_c$ along a direction induced by the transport map $Id + \xi$, where ξ is a smooth function with compact support such that $\text{supp}((Id + s\xi)_{\#}\nu) \subseteq \overline{B(0, R_T)}$ for any sufficiently small $s > 0$. If we denote with

$$
\begin{aligned}
\mathcal{L}_1(s) &= \ell(y, \pi_1(z_0) + s(\pi_1 \circ \xi)(z_1), \textstyle\int \omega d(\pi_1 \circ (Id + s\xi))_{\#}\nu), \\
\mathcal{L}_2(s) &= \ell(y, \pi_1(z_0), \textstyle\int \omega d(\pi_1 \circ (Id + s\xi))_{\#}\nu),
\end{aligned}
$$

then the map

$$
\begin{aligned}
s \quad \mapsto \quad & \frac{\mathcal{F}((z_0 - z_1) + s(\xi(z_0) - \xi(z_1))) - \mathcal{F}(z_0 - z_1)}{s} \\
& + \frac{\mathcal{G}(z_0 + s\xi(z_0)) - \mathcal{G}(z_0)}{s} - \frac{\mathcal{L}_1(s) - \mathcal{L}_2(s)}{s} - \frac{\mathcal{L}_2(s) - \mathcal{L}_2(0)}{s},
\end{aligned}
$$

as $s \to 0$ converges to

$$
\nabla \mathcal{F}(z_0 - z_1) \cdot (\xi(z_0) - \xi(z_1)) + (\nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot \xi(z_0).
$$

Since ν has compact support, the dominated convergence theorem, the identity (4.11) and since $\nabla \mathcal{F}$ is odd, it holds

$$
\begin{aligned}
+\infty \quad > \quad & \lim_{s \to 0} \frac{\mathbb{H}_c((Id + s\xi)_{\#}\nu) - \mathbb{H}_c(\nu)}{s} \\
= \quad & \frac{1}{2}\int_{\mathbb{R}^{4d}} \nabla \mathcal{F}(z_0 - z_1) \cdot (\xi(z_0) - \xi(z_1))d\nu(z_0)d\nu(z_1) + \int_{\mathbb{R}^{2d}} (\nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot \xi(z_0)d\nu(z_0) \\
= \quad & \int_{\mathbb{R}^{2d}} w(z_0) \cdot \xi(z_0)d\nu(z_0).
\end{aligned}
$$

From the last inequality, the assumption that $|\partial \mathbb{H}_c|(\nu) < +\infty$ and using the trivial estimate

$$
\mathcal{W}_2((Id + s\xi)_{\#}\nu, \nu) \le s\|\xi\|_{L^2_\nu},
$$

we get

$$
\int_{\mathbb{R}^{2d}} w(z_0) \cdot \xi(z_0)d\nu(z_0) \le |\partial \mathbb{H}_c|(\nu)\|\xi\|_{L^2_\nu},
$$

and hence, changing the sign of ξ,

$$
\left| \int_{\mathbb{R}^{2d}} w(z_0) \cdot \xi(z_0)d\nu(z_0) \right| \le |\partial \mathbb{H}_c|(\nu)\|\xi\|_{L^2_\nu}.
$$

This proves that $w \in L^2_\nu(\mathbb{R}^{2d})$ and that $\|w\|_{L^2_\nu} \le |\partial \mathbb{H}_c|(\nu)$.

We now prove that if the vector w belongs to $L^2_\nu(\mathbb{R}^{2d})$, then it is in the subdifferential of $\mathbb{H}_c$; this shall imply, by (4.1), that $w \in D(|\partial \mathbb{H}_c|)$ and that it is a minimal selection $\partial \mathbb{H}_c(\nu)$, by the previous estimate and Proposition 4.2.

We thus consider a test measure $\overline{\nu}$, a plan $\rho \in \Gamma_o(\nu, \overline{\nu})$, and we compute the directional derivative of $\mathbb{H}_c$ along the direction induced by ρ. Denoting with

$$
\begin{aligned}
\mathcal{L}_1(s) &= \ell(y, (1 - s)z_0 + sz_1, \textstyle\int \omega d((1 - s)\pi_1 + s\pi_2)_{\#}\rho), \\
\mathcal{L}_2(s) &= \ell(y, z_0, \textstyle\int \omega d((1 - s)\pi_1 + s\pi_2)_{\#}\rho),
\end{aligned}
$$

for every $s \in [0, 1]$, then the map

$$s \quad \mapsto \quad \frac{\mathcal{F}((1-s)(z_0 - \overline{z}_0) + s(z_1 - \overline{z}_1)) - \mathcal{F}(z_0 - \overline{z}_0)}{s}$$
$$+ \frac{\mathcal{G}((1-s)z_0 + sz_1) - \mathcal{G}(z_0)}{s} - \frac{\mathcal{L}_1(s) - \mathcal{L}_2(s)}{s} - \frac{\mathcal{L}_2(s) - \mathcal{L}_2(0)}{s},$$

as $s \to 0$ converges to

$$\nabla \mathcal{F}(z_0 - \overline{z}_0) \cdot ((z_1 - z_0) - (\overline{z}_1 - \overline{z}_0)) + (\nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot (z_1 - z_0).$$

Hence, from Proposition 4.10, the dominated convergence theorem, the identity (4.11) and since $\nabla \mathcal{F}$ is odd, we get

$$
\begin{aligned}
\mathbb{H}_c(\overline{\nu}) - \mathbb{H}_c(\nu) \quad &\geq \quad \lim_{s \to 0} \frac{\mathbb{H}_c(((1-s)\pi_1 + s\pi_2)_{\#}\rho) - \mathbb{H}_c(\nu)}{s} + o(\mathcal{W}_2(\overline{\nu}, \nu)) \\
&= \quad \frac{1}{2} \int_{\mathbb{R}^{8d}} \nabla \mathcal{F}(z_0 - \overline{z}_0) \cdot ((z_1 - z_0) - (\overline{z}_1 - \overline{z}_0)) \, d\rho(z_0, z_1) d\rho(\overline{z}_0, \overline{z}_1) \\
&\quad + \int_{\mathbb{R}^{4d}} (\nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot (z_1 - z_0) d\rho(z_0, z_1) + o(\mathcal{W}_2(\overline{\nu}, \nu)) \\
&= \quad \int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) + o(\mathcal{W}_2(\overline{\nu}, \nu)).
\end{aligned}
$$

We have thus proven that $w \in \partial \mathbb{H}_c(\nu)$. $\qquad \square$

5 Proof of Theorem 1.1

In this section, we prove Theorem 1.1. We first recall that we already proved in Corollary 2.15 that there exists a mean-field optimal control for Problem 1. We now want to prove that all mean-field optimal controls are solutions of the extended PMP.

Let u^* be a mean-field optimal control for Problem 1 with initial datum (y^0, μ^0). Fix μ_N^0 as in Definition 1.4–(i), and consider a sequence $(u_N^*)_{N \in \mathbb{N}}$ of optimal controls of Problem 2 with initial datum (y^0, μ_N^0), having a subsequence (which, for simplicity, we do not relabel) weakly converging to u^* in $L^1([0, T]; \mathcal{U})$. Denote with (y_N^*, x_N^*) the trajectory of (1.8) corresponding to the control u_N^* and the initial datum (y^0, μ_N^0) of Problem 2. Compute the corresponding pair control-trajectory $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying the PMP for Problem 2, that exists due to Theorem 3.2. Set $\nu_N^* := \Pi_N(x_N^*, p_N^*)$ and $r_N^* := Np_N^*$. By Proposition 3.5, the trajectories (y_N^*, q_N^*, ν_N^*) are equibounded and equi-Lipschitz from $[0, T]$ to the product space $\mathcal{Y} = \mathbb{R}^{2dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ endowed with the distance eqrefe-Y, and the empirical measures ν_N^* have equibounded support. Moreover, the pair $(u_N^*, (y_N^*, q_N^*, \nu_N^*))$ satisfies the extended PMP by Proposition 3.6.

By the Ascoli-Arzelà theorem, we have that there exists a subsequence, which we denote again with (y_N^*, q_N^*, ν_N^*), that converges to $(y^*, q^*, \nu^*) : [0, T] \to \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$ uniformly with respect to $t \in [0, T]$. Since by definition $\pi_{1\#}\nu_N^* = \mu_N^*$, by the convergence of μ_N^* to μ^* proved in Proposition 2.6, we get $\pi_{1\#}\nu^* = \mu^*$. Observe that (y^*, q^*, ν^*) is a Lipschitz function with respect to time and ν^* has support contained in $B(0, R_T)$ for all $t \in [0, T]$. Moreover, by the boundary conditions for each N, we have that $y^*(0) = y^0$, $\pi_{1\#}(\nu^*(0)) = \mu^0$ and $q^*(T) = 0$, $\pi_{2\#}(\nu^*(T))(r) = \delta(r)$.

Fix now $t \in [0, T]$. To shorten notation, let $E : \mathbb{R}^{dm} \times \mathbb{R}^D \to \mathbb{R}$ be the functional, strictly concave with respect to u, defined as

$$E(q, u) = \sum_{k=1}^m q_k \cdot B_k u - \gamma(u).$$

Recall that by (3.1) and by Remark 3.4, $u_N^*(t)$ satisfies

$$u_N^*(t) = \arg\max_{u \in \mathcal{U}} E(q_N^*(t), u),$$

since the maximum is uniquely determined by strict concavity. Since $\mathcal{U}$ is bounded, by definition $E(\cdot, u)$ is continuous uniformly with respect to $u \in \mathcal{U}$. The convergence of $q_N^*(t)$ to $q^*(t)$ then implies that every accumulation point $v_t \in \mathcal{U}$ of $u_N^*(t)$ must satisfy

$$v_t = \arg\max_{u \in \mathcal{U}} E(q^*(t), u) \tag{5.1}$$

and is therefore uniquely determined. This shows that the sequence u_N^* is pointwise converging in $[0, T]$ to the function $v(t) := v_t$. Due to the boundedness of $\mathcal{U}$, we further have that $u_N^* \to v$ in $L^1((0, T); \mathcal{U})$. Since u_N^* was already converging to u^* weakly in $L^1((0, T); \mathcal{U})$ it must be $u^*(t) = v(t)$ for a.e. $t \in (0, T)$, which together with (5.1) implies that

$$u_N^* \to u^* \text{ strongly in } L^1((0, T); \mathcal{U}) \tag{5.2}$$

and that

$$u^*(t) = \arg\max_{u \in \mathcal{U}} E(q^*(t), u)$$

for a.e. $t \in [0, T]$. Due to the explicit expression of $\mathbb{H}(y, q, \nu, u)$ in (1.4), this is equivalent to say that

$$\mathbb{H}(y^*(t), q^*(t), \nu^*(t), u^*(t)) = \arg\max_{u \in \mathcal{U}} \mathbb{H}(y^*(t), q^*(t), \nu^*(t), u)$$

for a.e. $t \in [0, T]$.

We finally prove that (y^*, q^*, ν^*) satisfies the Hamiltonian system (1.3) with control u^*. Due to equi-Lipschitz continuity, we have that the derivatives $(\dot{y}_N^*, \dot{q}_N^*)$, and $\partial_t \nu_N^*$ converge to $(\dot{y}^*, \dot{q}^*)$, and $\partial_t \nu^*$, respectively, weakly in $L^1([0, T]; \mathbb{R}^{2md})$ and in the sense of distributions. Observe now that by (1.5) and (1.6) the vector field $\nabla_\nu \mathbb{H}_c(y, q, \nu)(\cdot, \cdot)$, which is independent of u, is continuously depending on (y, q, ν). By the uniform convergence of (y_N^*, q_N^*, ν_N^*) and since $\text{supp}(\nu_N^*(t)) \subset B(0, R_T)$ for all $t \in [0, T]$ we get that

$$\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))(x, r) \rightrightarrows \nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t))(x, r)$$

uniformly with respect to $t \in [0, T]$ and $(x, r) \in B(0, R_T)$. From this, using again the narrow convergence of $\nu_N^*(t)$ to $\nu^*(t)$ and since $\text{supp}(\nu_N^*(t)) \subset B(0, R_T)$, we then get the uniform bound

$$\| (J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))) \nu_N^*(t) \|_{M_b(\mathbb{R}^D, \mathbb{R}^D)} \leq C_T,$$

for some constant C_T independent of $t \in [0, T]$, as well as the narrow convergence

$$(J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t))) \nu_N^*(t) \rightharpoonup (J\nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t))) \nu^*(t)$$

for all $t \in [0, T]$. Testing with functions $\phi \in \mathcal{C}_c^\infty([0, T] \times \mathbb{R}^{2d}; \mathbb{R})$, the two above properties are enough to show that

$$\nabla_{(x,r)} \cdot ((J\nabla_\nu \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t)))\nu_N^*(t)) \rightharpoonup \nabla_{(x,r)} \cdot ((J\nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t)))\nu^*(t))$$

in the sense of distributions, so that ν^* solves the third equation in (1.3).

For all $k = 1, \ldots, m$, taking derivatives in the explicit expression in (1.4) and using the definition of $\mathbb{H}_c$, we have that $\nabla_{y_k} \mathbb{H}_c(y, q, \nu, u)$ is actually independent of u and is continuous with respect to

the Euclidean convergence on (y, q) and the narrow convergence on measures ν with compact support in a fixed ball $B(0, R_T)$. Therefore, since (y_N^*, q_N^*, ν_N^*) converges to (y^*, q^*, ν^*) uniformly with respect to $t \in [0, T]$, and there is no dependence on u, for all $k = 1, \ldots, m$ we have that

$$\nabla_{y_k} \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) \to \nabla_{y_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t))$$

in $\mathbb{R}^d$ uniformly with respect to $t \in [0, T]$. It then follows that q^* solves the second equation in (1.3).

A similar argument, also using the L^1 convergence of u_N^* to u^* proved in (5.2), shows that

$$\nabla_{q_k} \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) \to \nabla_{q_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t))$$

in $L^1([0, T]; \mathbb{R}^d)$ for all $k = 1, \ldots, m$, so that y^* solves the first equation in (1.3). This concludes the proof of Theorem 1.1.

6 An example

In this section, we show the application of the extended Pontryagin Maximum Principle to a toy model for crowd interactions. The Cucker-Smale model, introduced in [18], was first studied in its mean-field limit form in [24]. It models the phenomenon of alignment of velocities in crowds, that can be observed, e.g., in flocks of birds.

In this model, each agent is identified by its position x_i and velocity v_i, and it adjusts its velocity by relaxing it towards a weighted mean of the velocities of the group. The weight is a nonincreasing function ϕ of the distance between individuals. In the original paper [18], the authors propose $\phi(\lambda) = \frac{K}{(\sigma^2 + \lambda^2)^\beta}$, for some fixed parameters $K, \sigma > 0$ and $\beta \geq 0$. For our computations, we consider $\phi \in \mathcal{C}^2(\mathbb{R}^d, \mathbb{R}^+)$ being a radial function.

The finite-dimensional dynamics is given by the ODE system

$$\begin{cases} \dot{x}_i = v_i, \\ \dot{v}_i = \dfrac{1}{N} \sum_{j=1}^{N} \phi(x_i - x_j)(v_j - v_i), \end{cases} \quad i = 1, \ldots, N.$$

We add to it m leaders with positions and velocities given by (y_k, w_k) for every $k = 1, \ldots, m$, on which a control variable u_k is active. Since the control acts as an external force, u_k will directly affect the evolution of the velocities w_k only. The mean-field limit for $N \to +\infty$ of the resulting system is given by (see, e.g., [22])

$$\begin{cases} \dot{y}_k = w_k, \\ \dot{w}_k = (\Phi \star \mu)(y_k, w_k) + \dfrac{1}{m} \sum_{j=1}^{m} \phi(y_k - y_j)(w_j - w_k) + u_k, \quad k = 1, \ldots, m \\ \partial_t \mu = -v \cdot \nabla_x \mu - \nabla_v \cdot \left[\left(\Phi \star \mu + \dfrac{1}{m} \sum_{j=1}^{m} \phi(x - y_j)(w_j - v) \right) \mu \right], \end{cases} \quad (6.1)$$

where $\mu = \mu(x, v)$ is the density of followers and $\Phi(x, v) := \phi(x)(-v)$. Notice, this is a particular case of (1.2), where the state variables for the leaders are $\mathbf{y}_k := \begin{pmatrix} y_k \\ w_k \end{pmatrix}$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, the

ones for the followers are $\mathbf{x} = \begin{pmatrix} x \\ v \end{pmatrix}$, and one chooses

$$K(\mathbf{x}) := \begin{pmatrix} 0 \\ \Phi(\mathbf{x}) \end{pmatrix}, \quad f_k(\mathbf{y}) = \begin{pmatrix} w_k \\ \frac{1}{m} \sum_{j=1}^m \phi(y_k - y_j)(w_j - w_k) \end{pmatrix},$$

$$g(\mathbf{y})(\mathbf{x}) = \begin{pmatrix} v \\ \frac{1}{m} \sum_{j=1}^m \phi(x - y_j)(w_j - v) \end{pmatrix},$$

and, for every $k = 1, \ldots, m$, B_k is the $2d \times (dm)$ matrix that maps $u = \begin{pmatrix} u_1 \\ \ldots \\ u_m \end{pmatrix} \in \mathbb{R}^{dm}$ into the

element $\begin{pmatrix} \mathbf{0} \\ u_k \end{pmatrix} \in \mathbb{R}^{2d}$. Notice that since ϕ is a radial function, the function Φ, and thus K, is odd.

A standard problem in the study of the Cucker-Smale model is to find conditions to ensure flocking, i.e., alignment of the whole crowd to the same velocity. For this reason, it is interesting in our case to study the minimization of the variance[2] of the crowd, by choosing

$$L(\mathbf{y}, \mu) := \int_{\mathbb{R}^{2d}} \left(\frac{2}{m} \sum_{k=1}^m \|w_k\|^2 + 2\|v\|^2 \right) d\mu(x, v) - \left\| \frac{1}{m} \sum_{k=1}^m w_k + \int v \, d\mu(x, v) \right\|^2$$

$$= \int_{\mathbb{R}^{2d}} \left(\frac{2}{m} \sum_{k=1}^m \|w_k\|^2 + 2\|v\|^2 - \left(\frac{1}{m} \sum_{k=1}^m w_k + \int v' \, d\mu(x', v') \right) \cdot \left(\frac{1}{m} \sum_{k=1}^m w_k + v \right) \right) d\mu(x, v),$$

(6.2)

that is of the form $L = \int_{\mathbb{R}^{2d}} \ell(\mathbf{y}, \mathbf{x}, \int \omega\mu) \, d\mu(\mathbf{x})$ by choosing $\omega(\mathbf{x}) = v$ and

$$\ell(\mathbf{y}, \mathbf{x}, \varsigma) = \frac{2}{m} \sum_{k=1}^m \|w_k\|^2 + 2\|v\|^2 - \left(\frac{1}{m} \sum_{k=1}^m w_k + \varsigma \right) \cdot \left(\frac{1}{m} \sum_{k=1}^m w_k + v \right).$$

For the control constraints, we assume $\mathcal{U} := [-1, 1]^{dm}$ and we choose to penalize the L^2-norm of the control, hence $\gamma(u) := \|u\|^2$.

Remark 6.1. Other forms for the cost L can be of interest. For example, on may want to drive the crowd to a *given* fixed velocity $\bar{v}$. In this case, one can minimize

$$L_1(\mathbf{y}, \mu) := \int_{\mathbb{R}^{2d}} \left(\frac{1}{2m} \sum_{k=1}^m \|w_k - \bar{v}\|^2 + \frac{1}{2}\|v - \bar{v}\|^2 \right) d\mu(x, v),$$

that is again of the form $\int_{\mathbb{R}^{2d}} \ell(\mathbf{y}, \mathbf{x}, \int \omega\mu) \, d\mu(\mathbf{x})$, with ℓ not depending on its third variable, this time.

Since Hypotheses (H) are clearly satisfied, we now apply the extended Pontryagin Maximum Principle to the optimal control problem with cost functional (6.2) constrained by the system (6.1). For simplicity of notation, we study the 1-dimensional problem, i.e., $d = 1$. We introduce the dual variables of $\mathbf{y}_k$ and $\mathbf{x}$ denoted by $\mathbf{q}_k = (q_k, z_k)$ and $\mathbf{r} = (r, s)$, respectively. The Hamiltonian $\mathbb{H}$ in

---

[2]For simplicity of computation, we consider minimization of 4 times the variance.

(1.4) can be found by direct substitution:

$$\mathbb{H}(\mathbf{y}, \mathbf{q}, \nu, u) = \frac{1}{2} \int_{\mathbb{R}^8} (s - s')\phi(x - x')(v' - v) \, d\nu(x', v', r', s') \, d\nu(x, v, r, s)$$

$$+ \int_{\mathbb{R}^4} \left(rv + s\frac{1}{m} \sum_{j=1}^{m} \phi(x - y_j)(w_j - v) \right) d\nu(x, v, r, s)$$

$$+ \sum_{k=1}^{m} \left(q_k w_k + z_k \int_{\mathbb{R}^4} \phi(y_k - x)(v - w_k) \, d\nu(x, v, r, s) \right)$$

$$+ \sum_{k=1}^{m} \left(z_k \frac{1}{m} \sum_{j=1}^{m} \phi(y_k - y_j)(w_j - w_k) + z_k u_k \right)$$

$$- \int_{\mathbb{R}^2} \ell(\mathbf{y}, \mathbf{x}, \int \omega\mu) \, d\mu(\mathbf{x}) - |u|^2.$$

The optimal control can be explicitly computed by (3.2) as follows

$$u_k^*(\mathbf{y}, \mathbf{q}) = \begin{cases} \frac{1}{2}z_k & \text{if } z_k \in [-2, 2], \\ \text{sign}(z_k) & \text{otherwise.} \end{cases}$$

Denoting with μ the first marginal of ν, the PMP dynamics of the state and adjoint variables is given by

$$\begin{cases} \dot{y}_k = & w_k, \\ \dot{w}_k = & (\Phi \star \mu)(y_k, w_k) + \frac{1}{m} \sum_{k=1}^{m} \phi(y_k - y_j)(v_j - v_k) + u_k^*(\mathbf{y}, \mathbf{q}), \\ \dot{q}_k = & \frac{1}{m} \int_{\mathbb{R}^4} s\phi'(x - y_k)(w_k - v) \, d\nu(x, v, r, s) - z_k \int_{\mathbb{R}^4} \phi'(y_k - x)(v - w_k) \, d\nu(x, v, r, s) \\ & -\frac{1}{m} \sum_{j \neq k} z_j \phi'(y_k - y_j)(w_j - w_k), \\ \dot{z}_k = & -\int_{\mathbb{R}^4} \left(\frac{1}{m}s - z_k \right) \phi(x - y_k) \, d\nu(x, v, r, s) - q_k + \sum_{j \neq k} z_j \phi(y_k - y_j) \\ & -\frac{4}{m} w_k + \frac{2}{m^2} \sum_{j=1}^{m} w_j + \frac{2}{m} \int_{\mathbb{R}^2} v \, d\mu(x, v), \\ \partial_t \nu = & -\nabla_{(x,v,r,s)} \cdot ((J\nabla_\nu \mathbb{H}_c(\mathbf{y}, \mathbf{q}, \nu, u^*))\nu), \end{cases}$$

where the components of the vector field $\nabla_\nu \mathbb{H}_c(\mathbf{y}, \mathbf{q}, \nu, u^*)$ are given at every point $(x, v, r, s) \in \mathbb{R}^4$ by

$$\nabla_\nu \mathbb{H}_c \cdot e_1 = \int_{\mathbb{R}^4} (s - s')\phi'(x - x')(v' - v) \, d\nu(x', v', r', s') + s\frac{1}{m} \sum_{j=1}^{m} \phi'(x - y_j)(w_j - v)$$

$$- \sum_{k=1}^{m} z_k \phi'(y_k - x)(v - w_k),$$

$$\nabla_\nu \mathbb{H}_c \cdot e_2 = - \int_{\mathbb{R}^4} (s - s')\phi(x - x') \, d\nu(x', v', r', s') + r - s\frac{1}{m} \sum_{j=1}^{m} \phi(x - y_j) + \sum_{k=1}^{m} z_k \phi(y_k - x) - 3v$$

$$+ \frac{2}{m} \sum_{k=1}^{m} w_k + \int_{\mathbb{R}^2} v d\mu(x, v),$$

$$\nabla_\nu \mathbb{H}_c \cdot e_3 = v,$$

$$\nabla_\nu \mathbb{H}_c \cdot e_4 = (\Phi \star \mu)(x, v) + \frac{1}{m} \sum_{j=1}^{m} \phi(x - y_j)(w_j - v).$$

We remark that, as it happens for the standard PMP, the explicit computation of the third and the fourth components gives exactly the vector field determining the dynamics of μ, the first marginal of ν, in accordance with (6.1).

Acknowledgement

References

[1] G. Albi, M. Bongini, E. Cristiani, and D. Kalise. Invisible sparse control of self-organizing agents leaving unknown environments. *Submitted*, 2015.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.*, volume 254. Clarendon Press Oxford, 2000.

[3] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

[4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[5] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Appl. Math. Comput.*, 63(3):341–356, 2011.

[6] A. Bensoussan, J. Frehse, and P. Yam. *Mean Field Games and Mean Field Type Control Theory.* Springer, 2013.

[7] M. Bongini and M. Fornasier. Sparse stabilization of dynamical systems driven by attraction and avoidance forces. *Netw. Heterog. Media*, 9(1):1–31, 2014.

[8] M. Bongini, M. Fornasier, F. Frölich, and L. Hagverdi. Sparse control of force field dynamics. In *International Conference on NETwork Games, COntrol and OPtimization*, October 2014.

[9] M. Bongini, M. Fornasier, O. Junge, and B. Scharf. Sparse control of alignment models in high dimension. *To appear in Netw. Heterog. Media*, 2014. Preprint at arXiv:1408.6362.

[10] M. Bongini, M. Fornasier, and D. Kalise. (Un)conditional consensus emergence under feedback controls. *To appear in Discrete Contin. Dyn. Syst.*, 35(9), 2015. Preprint at arXiv:1502.06100.

[11] M. Burger, M. Di Francesco, P. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete Contin. Dyn. Syst. Ser. B*, 19:1311–1333, 2014.

[12] G. Buttazzo and G. D. Maso. Γ-convergence and optimal control problems. *J. Optim. Theory Appl.*, 38(3):385–407, 1982.

[13] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat. Sparse stabilization and control of alignment models. *Math. Models Methods Appl. Sci.*, 25(03):521–564, 2015.

[14] R. Carmona, F. Delarue, and A. Lachapelle. Control of McKean–Vlasov dynamics versus mean field games. *Math. Financ. Econ.*, 7(2):131–166, 2013.

[15] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*, volume 264. Springer, 2013.

[16] C. Clason and K. Kunisch. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM Control Optim. Calc. Var.*, 17(01):243–266, 2011.

[17] F. Cucker and J.-G. Dong. A general collision-avoiding flocking framework. *IEEE Trans. Automat. Control*, 56(5):1124–1129, 2011.

[18] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862, 2007.

[19] M. Cullen, W. Gangbo, and G. Pisante. The semigeostrophic equations discretized in reference and dual variables. *Archive for Rational Mechanics and Analysis*, 185(2):341–363, 2007.

[20] G. Dal Maso. *An Introduction to Γ-Convergence*. Progr. Nonlinear Differential Equations Appl. Birkhäuser Boston Inc., Boston, MA, 1993.

[21] M. R. D'Orsogna, Y.-L. Chuang, A. L. Bertozzi, and L. S. Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Phys. Rev. Lett.*, 96(10):104302, 2006.

[22] M. Fornasier, B. Piccoli, and F. Rossi. Mean-field sparse optimal control. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130400, 2014.

[23] M. Fornasier and F. Solombrino. Mean-field optimal control. *ESAIM Control Optim. Calc. Var.*, 20(4):1123–1152, 2014.

[24] S.-Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic description of flocking. *Kinet. Relat. Models*, 1(3):415–435, 2008.

[25] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[26] M. Huang, P. E. Caines, and R. P. Malhamé. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and nash equilibrium solutions. In *42nd IEEE Conference on Decision and Control*, volume 1, pages 98–103. IEEE, 2003.

[27] R. L. Hughes. A continuum theory for the flow of pedestrians. *Transport. Res. B Meth.*, 36(6):507–535, 2002.

[28] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and their Applications*. Academic Press, New York, NY, 1980.

[29] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[30] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Rev.*, 56(4):577–621, 2014.

[31] B. Piccoli and F. Rossi. Transport equation with nonlocal velocity in Wasserstein spaces: Convergence of numerical schemes. *Acta Appl. Math.*, 124(1):73–105, 2013.

[32] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, 211(1):335–358, 2014.

[33] B. Piccoli, F. Rossi, and E. Trélat. Control to flocking of the kinetic cucker-smale model. *Submitted*, 2014. Preprint at arXiv:1411.4687.

[34] K. Pieper and B. Vexler. A priori error analysis for discretization of sparse elliptic optimal control problems in measure space. *SIAM J. Control Optim.*, 51(4):2788–2808, 2013.

[35] Y. Privat, E. Trélat, and E. Zuazua. Optimal location of controllers for the one-dimensional wave equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 30(6):1097–1126, 2013.

[36] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Appl. Math. Optim.*, 39(2):143–177, 1999.

[37] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4):25–34, 1987.

[38] G. Stadler. Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. *Comput. Optim. Appl.*, 44(2):159–181, 2009.

[39] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75(6):1226, 1995.

[40] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.